VARIABILITY STRUCTURE ESTIMATION: THE ROLE OF SAMPLE SUPPORT

BY

MIGUEL RESTREPO

Ingen., Universidad de Los Andes, Bogotá, 1991
M.E., Cornell University, 1993
M.S., University of Illinois, 1996

THESIS

Submitted in partial fullfilment of the requirements
for the degree of Doctor of Philosophy in Natural Resources and Environmental Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2000

Urbana, Illinois

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

THE GRADUATE COLLEGE

OCTOBER 2000
(date)

WE HEREBY RECOMMEND THAT THE THESIS BY

MIGUEL RESTREPO

ENTITLED VARIABILITY STRUCTURE ESTIMATION:

THE ROLE OF SAMPLE SUPPORT

BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF DOCTOR OF PHILOSOPHY

Director of Thesis Research

Head of Department

Committee on Final Examination†

Chairperson

† Required for doctor's degree but not for master's.

O-517

UNIVERSITY OF ILLINOIS
AT URBANA-CHAMPAIGN

GRADUATE COLLEGE
DEPARTMENTAL FORMAT APPROVAL

THIS IS TO CERTIFY THAT THE FORMAT AND QUALITY OF PRESENTATION OF THE THESIS SUBMITTED BY

MIGUEL RESTREPO

AS ONE OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

ARE ACCEPTABLE TO THE

DEPARTMENT OF NATURAL RESOURCES AND ENVIRONMENTAL SCIENCES
Full Name of Department, Division, or Unit

7/12/00
Date of Approval

John A. Juick
Departmental Representative

VARIABILITY STRUCTURE ESTIMATION: THE ROLE OF SAMPLE SUPPORT

Miguel Restrepo, Ph.D.
Department of Natural Resources and Environmental Sciences
University of Illinois at Urbana-Champaign, 2000
Timothy Ellsworth, Advisor

The characterization of spatial variability in soil properties is a prerrequisite to many activities, such as site specific management (i.e., precision farming), groundwater and solute transport modeling, groundwater pollution assessment and remediation, etc. Due to soil heterogeneity, statistical measures are often used for variability description. The experimental point variogram and the experimental histogram are the two most widely used statistical measures of variability structure. For a given spatial domain, it is often difficult to accurately estimate the point variogram due to sampling costs and limited resources, thus we aim to maximize the relative information in our data sampling efforts. The work presented in this document shows in detail the development of a methodology to estimate the point variogram using different types of regularized data, i.e., single-support variograms and mixed-support variograms. The applicability of the method is shown using two different sets of data one is a conditional simulation based on 1650 measurements of phosphorous in a section of one mi^2 area (640 acres, or 259 ha). The second data set consists of chloride mass recovery measurements within a small field plot. With these application examples, the relative information content of different measurement methods for characterizing the point variogram are evaluated (in terms of integral scale and sill). The effectiveness and robustness of the methodology are analyzed by means of Monte Carlo analyses, using the conditional simulation data. A cation exchange capacity (CEC) field was obtained by a conditional simulation of 1650 CEC measurements from from Williams Field and is used with several sampling schemes to analyze the influence of sampling patterns in three different numerical experiments. The results are shown as confidence intervals of the estimated variogram, the histogram of the data, and the spread of the confidence intervals of the parameters obtained from the data fitting routines.

Acknowledgements

I would like to thank my parents, for the constant and unconditional love and support that they have always generously given. I also want to thank my advisor and friend, Professor Tim Ellsworth for his generosity and helpful guidance and comments during my studies and the preparation of this thesis. I want to express special gratitude to the Illinois State Water Survey during my last stage of the program at the University. In particular, I want to mention my supervisors, Dr. David Soong and Mrs. Laura Keefer for their help and support during difficult economic times. Also, thank you to the other students and employees that I somehow worked with: Amy Russell, Dan Kriesant, Tim Grimm, Selena Worster, Kristin Jaburek, Megan Stewart, Yanqing Lian and all the others that I forgot to mention. Their company and moral support was always important at the time of "the big push" to complete my thesis. And, finally, thank you to all my friends and collegues who always kept an interest and criticism regarding my ideas and objectives during the program; it is always good to have a healthy and more objective point of view of reality.

Contents

Chapter 1 Introduction, Literature Background, and Objectives **1**

1.1 Introduction 1

1.2 Literature Review 3

1.3 Objectives 55

Chapter 2 Theory and Methodology **58**

2.1 Assumptions 58

2.2 Concepts and Definitions 59

2.3 Experimental Variograms and Dispersion Variances 61

2.4 Numerical Approximation of Regularized Variograms and Dispersion Variances 65

Chapter 3 Variability Structure Estimation **67**

3.1 Explicit Relationship of γ_v, $\bar{\gamma}$, and γ 67

3.2 Error Minimization Problem for Variogram Deconvolution 68

3.3 Nested Structures for Nugget Analysis 72

Chapter 4 Applications of Variogram Deconvolution **74**

4.1 Field Measured Data: Cores + Donuts 74

4.2 Conditional Simulations 76

4.2.1 Deciding How to Generate the CEC $1024x1024$ Data Sets 79

4.2.2 Williams Field: Conditional Simulation for Phosphorous 85

Chapter 5 Robustness of the Method **89**

5.1 Monte Carlo Methods for Confidence Intervals Computation 90

5.2 Histogram Transformations for Change of Sample Support 93

5.3 Numerical Sampling Experiments: Monte Carlo applications 96

5.3.1 Numerical Experiment Number One 98

5.3.2 Numerical Experiment Number Two 100

5.3.3 Numerical Experiment Number Three 101

Chapter 6 Analysis of Results and Conclusions **105**

Appendix A Numerical Expressions for $\bar{\gamma}(h, v, V)$ **116**

A.1 1-Dimensional case 116

A.2 2-Dimensional case 119

A.3 3-Dimensional case 120

Appendix B Numerical Experiment 1 - Plain **123**

Appendix C Numerical Experiment 1 - Nugget **188**

Appendix D Numerical Experiments 2 and 3 - Plain **253**

Appendix E Numerical Experiments 2 and 3 - Nugget **276**

Bibliography **299**

Vita **307**

List of Tables

6.1 Experiment 1, Plain. Table entries are the values of $sill \cdot range$. Theoretical value is 402.47 112

6.2 Experiment 1, Nugget. Table entries are the values of $sill \cdot range$. Theoretical value is 353.76 113

6.3 Experiments 2 and 3, Plain. Table entries are the values of $sill \cdot range$. Theoretical value is 402.47 114

6.4 Experiments 2 and 3, Nugget. Table entries are the values of $sill \cdot range$. Theoretical value is 353.76 115

B.1 Experiment 1, Plain 12 x 12, sill = 7.223, range = 3*55.721 124

B.2 Experiment 1, Plain, 16 x 16, sill = 7.223, range = 3*55.721 133

B.3 Experiment 1, Plain, 24 x 24, sill = 7.223, range = 3*55.721 142

B.4 Experiment 1, Plain, 32 x 32, sill = 7.223, range = 3*55.721 151

B.5 Experiment 1, Plain, 40 x 40, sill = 7.223, range = 3*55.721 160

C.1 Experiment 1, Nugget, 12 x 12, sill = 5.689, range = 3*62.184 189

C.2 Experiment 1, Nugget, 16 x 16, sill = 5.689, range = 3*62.184 198

C.3 Experiment 1, Nugget, 24 x 24, sill = 5.689, range = 3*62.184 207

C.4 Experiment 1, Nugget, 32 x 32, sill = 5.689, range = 3*62.184 216

C.5 Experiment 1, Nugget, 40 x 40, sill = 5.689, range = 3*62.184 225

D.1 Experiment 2 Plain, 24x24, Sill = 7.223, Range = 3*55.721 254

D.2 Experiment 3 Plain, 12x12, Sill = 7.223, Range = 3*55.721 265

E.1 Experiment 2 Nugget, 24x24, Sill = 5.689, Range = 3*62.184 277

E.2 Experiment 3 Nugget, 12x12, Sill = 5.689, Range = 3*62.184 288

List of Figures

4.1 Sampling layout of "cores" and "donuts" 75

4.2 Regularized and point variograms for "cores" and "donuts" 76

4.3 Observed and computed dispersion variances for "cores" and "donuts" 77

4.4 BIGAUS results compared to Indicator variograms for CEC 79

4.5 Variogram Results and Model fit for CEC Plain Conditional Simulation . . . 81

4.6 Variogram Results and Model fit for CEC Nugget Conditional Simulation . . 82

4.7 Variogram Results and Model fit for CEC Original Observed Values 83

4.8 q-q plot of observed CEC values and CEC Plain Conditional Simulation Results 84

4.9 q-q plot of observed CEC values and CEC Nugget Conditional Simulation Results 85

4.10 Sampling layout at Williams Field for Phosphorous Example 86

4.11 Regularized variograms and fitted point variogram for Phosphorous Example 87

4.12 Experimental and Fitted Dispersion Variances for Phosphorous Example . . 88

5.1 Typical realization Experiment 2: 24x24 102

5.2 Typical realization Experiment 3: 12x12 104

A.1 1-Dimensional mixed support overlaping length 117

A.2 2-Dimensional mixed support overlaping area 119

B.1 Numerical Experiment 1 Plain, 12x12 - Obs. Point 125

B.2 Numerical Experiment 1 Plain, 12x12 - Fit. Point 126

B.3 Numerical Experiment 1 Plain, 12x12 - L/80 127

B.4 Numerical Experiment 1 Plain, 12x12 - L/40 128
B.5 Numerical Experiment 1 Plain, 12x12 - mix 129
B.6 Numerical Experiment 1 Plain, 12x12 - point + L/80 130
B.7 Numerical Experiment 1 Plain, 12x12 - point + L/80 + L/40 131
B.8 Numerical Experiment 1 Plain, 12x12 - point + L/80 + L/40 + mix 132
B.9 Numerical Experiment 1 Plain, 16x16 - Obs. Point 134
B.10 Numerical Experiment 1 Plain, 16x16 - Fit. Point 135
B.11 Numerical Experiment 1 Plain, 16x16 - L/80 136
B.12 Numerical Experiment 1 Plain, 16x16 - L/40 137
B.13 Numerical Experiment 1 Plain, 16x16 - mix 138
B.14 Numerical Experiment 1 Plain, 16x16 - point + L/80 139
B.15 Numerical Experiment 1 Plain, 16x16 - point + L/80 + L/40 140
B.16 Numerical Experiment 1 Plain, 16x16 - point + L/80 + L/40 + mix 141
B.17 Numerical Experiment 1 Plain, 24x24 - Obs. Point 143
B.18 Numerical Experiment 1 Plain, 24x24 - Fit. Point 144
B.19 Numerical Experiment 1 Plain, 24x24 - L/80 145
B.20 Numerical Experiment 1 Plain, 24x24 - L/40 146
B.21 Numerical Experiment 1 Plain, 24x24 - mix 147
B.22 Numerical Experiment 1 Plain, 24x24 - point + L/80 148
B.23 Numerical Experiment 1 Plain, 24x24 - point + L/80 + L/40 149
B.24 Numerical Experiment 1 Plain, 24x24 - point + L/80 + L/40 + mix 150
B.25 Numerical Experiment 1 Plain, 32x32 - Obs. Point 152
B.26 Numerical Experiment 1 Plain, 32x32 - Fit. Point 153
B.27 Numerical Experiment 1 Plain, 32x32 - L/80 154
B.28 Numerical Experiment 1 Plain, 32x32 - L/40 155
B.29 Numerical Experiment 1 Plain, 32x32 - mix 156
B.30 Numerical Experiment 1 Plain, 32x32 - point + L/80 157

B.31 Numerical Experiment 1 Plain, 32x32 - point + L/80 + L/40 158
B.32 Numerical Experiment 1 Plain, 32x32 - point + L/80 + L/40 + mix 159
B.33 Numerical Experiment 1 Plain, 40x40 - Obs. point 161
B.34 Numerical Experiment 1 Plain, 40x40 - Fit. point 162
B.35 Numerical Experiment 1 Plain, 40x40 - L/80 163
B.36 Numerical Experiment 1 Plain, 40x40 - point + L/80 164
B.37 Histogram Experiment 1 Plain, 12x12 - Obs. point 165
B.38 Histogram Experiment 1 Plain, 12x12 - Afine L/80 166
B.39 Histogram Experiment 1 Plain, 12x12 - Log L/80 167
B.40 Histogram Experiment 1 Plain, 12x12 - Afine L/40 168
B.41 Histogram Experiment 1 Plain, 12x12 - Log L/40 169
B.42 Histogram Experiment 1 Plain, 16x16 - Obs. point 170
B.43 Histogram Experiment 1 Plain, 16x16 - Afine L/80 171
B.44 Histogram Experiment 1 Plain, 16x16 - Log L/80 172
B.45 Histogram Experiment 1 Plain, 16x16 - Afine L/40 173
B.46 Histogram Experiment 1 Plain, 16x16 - Log L/40 174
B.47 Histogram Experiment 1 Plain, 24x24 - Obs. point 175
B.48 Histogram Experiment 1 Plain, 24x24 - Afine L/80 176
B.49 Histogram Experiment 1 Plain, 24x24 - Log L/80 177
B.50 Histogram Experiment 1 Plain, 24x24 - Afine L/40 178
B.51 Histogram Experiment 1 Plain, 24x24 - Log L/40 179
B.52 Histogram Experiment 1 Plain, 32x32 - Obs. point 180
B.53 Histogram Experiment 1 Plain, 32x32 - Afine L/80 181
B.54 Histogram Experiment 1 Plain, 32x32 - Log L/80 182
B.55 Histogram Experiment 1 Plain, 32x32 - Afine L/40 183
B.56 Histogram Experiment 1 Plain, 32x32 - Log L/40 184
B.57 Histogram Experiment 1 Plain, 40x40 - Obs. point 185

B.58 Histogram Experiment 1 Plain, 40x40 - Afine L/80 186
B.59 Histogram Experiment 1 Plain, 40x40 - Log L/80 187
C.1 Numerical Experiment 1 Nugget, 12x12 - Obs. Point 190
C.2 Numerical Experiment 1 Nugget, 12x12 - Fit. Point 191
C.3 Numerical Experiment 1 Nugget, 12x12 - L/80 192
C.4 Numerical Experiment 1 Nugget, 12x12 - L/40 193
C.5 Numerical Experiment 1 Nugget, 12x12 - mix 194
C.6 Numerical Experiment 1 Nugget, 12x12 - point + L/80 195
C.7 Numerical Experiment 1 Nugget, 12x12 - point + L/80 + L/40 196
C.8 Numerical Experiment 1 Nugget, 12x12 - point + L/80 + L/40 + mix 197
C.9 Numerical Experiment 1 Nugget, 16x16 - Obs. Point 199
C.10 Numerical Experiment 1 Nugget, 16x16 - Fit. Point 200
C.11 Numerical Experiment 1 Nugget, 16x16 - L/80 201
C.12 Numerical Experiment 1 Nugget, 16x16 - L/40 202
C.13 Numerical Experiment 1 Nugget, 16x16 - mix 203
C.14 Numerical Experiment 1 Nugget, 16x16 - point + L/80 204
C.15 Numerical Experiment 1 Nugget, 16x16 - point + L/80 + L/40 205
C.16 Numerical Experiment 1 Nugget, 16x16 - point + L/80 + L/40 + mix 206
C.17 Numerical Experiment 1 Nugget, 24x24 - Obs. Point 208
C.18 Numerical Experiment 1 Nugget, 24x24 - Fit. Point 209
C.19 Numerical Experiment 1 Nugget, 24x24 - L/80 210
C.20 Numerical Experiment 1 Nugget, 24x24 - L/40 211
C.21 Numerical Experiment 1 Nugget, 24x24 - mix 212
C.22 Numerical Experiment 1 Nugget, 24x24 - point + L/80 213
C.23 Numerical Experiment 1 Nugget, 24x24 - point + L/80 + L/40 214
C.24 Numerical Experiment 1 Nugget, 24x24 - point + L/80 + L/40 + mix 215
C.25 Numerical Experiment 1 Nugget, 32x32 - Obs. Point 217

C.26 Numerical Experiment 1 Nugget, 32x32 - Fit. Point 218
C.27 Numerical Experiment 1 Nugget, 32x32 - L/80 219
C.28 Numerical Experiment 1 Nugget, 32x32 - L/40 220
C.29 Numerical Experiment 1 Nugget, 32x32 - mix 221
C.30 Numerical Experiment 1 Nugget, 32x32 - point + L/80 222
C.31 Numerical Experiment 1 Nugget, 32x32 - point + L/80 + L/40 223
C.32 Numerical Experiment 1 Nugget, 32x32 - point + L/80 + L/40 + mix 224
C.33 Numerical Experiment 1 Nugget, 40x40 - Obs. point 226
C.34 Numerical Experiment 1 Nugget, 40x40 - Fit. point 227
C.35 Numerical Experiment 1 Nugget, 40x40 - L/80 228
C.36 Numerical Experiment 1 Nugget, 40x40 - point + L/80 229
C.37 Histogram Experiment 1 Nugget, 12x12 - Obs. point 230
C.38 Histogram Experiment 1 Nugget, 12x12 - Afine L/80 231
C.39 Histogram Experiment 1 Nugget, 12x12 - Log L/80 232
C.40 Histogram Experiment 1 Nugget, 12x12 - Afine L/40 233
C.41 Histogram Experiment 1 Nugget, 12x12 - Log L/40 234
C.42 Histogram Experiment 1 Nugget, 16x16 - Obs. point 235
C.43 Histogram Experiment 1 Nugget, 16x16 - Afine L/80 236
C.44 Histogram Experiment 1 Nugget, 16x16 - Log L/80 237
C.45 Histogram Experiment 1 Nugget, 16x16 - Afine L/40 238
C.46 Histogram Experiment 1 Nugget, 16x16 - Log L/40 239
C.47 Histogram Experiment 1 Nugget, 24x24 - Obs. point 240
C.48 Histogram Experiment 1 Nugget, 24x24 - Afine L/80 241
C.49 Histogram Experiment 1 Nugget, 24x24 - Log L/80 242
C.50 Histogram Experiment 1 Nugget, 24x24 - Afine L/40 243
C.51 Histogram Experiment 1 Nugget, 24x24 - Log L/40 244
C.52 Histogram Experiment 1 Nugget, 32x32 - Obs. point 245

C.53 Histogram Experiment 1 Nugget, 32x32 - Afine L/80 246
C.54 Histogram Experiment 1 Nugget, 32x32 - Log L/80 247
C.55 Histogram Experiment 1 Nugget, 32x32 - Afine L/40 248
C.56 Histogram Experiment 1 Nugget, 32x32 - Log L/40 249
C.57 Histogram Experiment 1 Nugget, 40x40 - Obs. point 250
C.58 Histogram Experiment 1 Nugget, 40x40 - Afine L/80 251
C.59 Histogram Experiment 1 Nugget, 40x40 - Log L/80 252
D.1 Numerical Experiment 2 Plain, (192 points) - Point 255
D.2 Numerical Experiment 2 Plain, (192 points) - L/80 256
D.3 Numerical Experiment 2 Plain, (192 points) - L/40 257
D.4 Numerical Experiment 2 Plain, (192 points) - mixed point and L/80 258
D.5 Numerical Experiment 2 Plain, (192 points) - mixed point and L/40 259
D.6 Numerical Experiment 2 Plain, (192 points) - mixed L/80 and L/40 260
D.7 Numerical Experiment 2 Plain, (192 points) - point and L/80 261
D.8 Numerical Experiment 2 Plain, (192 points) - point and L/40 262
D.9 Numerical Experiment 2 Plain, (192 points) - point and L/80 and L/40 . . . 263
D.10 Numerical Experiment 2 Plain, (192 points) - all variograms at once 264
D.11 Numerical Experiment 3 Plain, (192 points) - point 266
D.12 Numerical Experiment 3 Plain, (192 points) - L/80 267
D.13 Numerical Experiment 3 Plain, (192 points) - L/40 268
D.14 Numerical Experiment 3 Plain, (192 points) - mixed point and L/80 269
D.15 Numerical Experiment 3 Plain, (192 points) - mixed point and L/40 270
D.16 Numerical Experiment 3 Plain, (192 points) - mixed L/80 and L/40 271
D.17 Numerical Experiment 3 Plain, (192 points) - point and L/80 272
D.18 Numerical Experiment 3 Plain, (192 points) - point and L/40 273
D.19 Numerical Experiment 3 Plain, (192 points) - point and L/80 and L/40 . . . 274
D.20 Numerical Experiment 3 Plain, (192 points) - all variograms at once 275

E.1 Numerical Experiment 2 Nugget, (192 points) - Point 278
E.2 Numerical Experiment 2 Nugget, (192 points) - L/80 279
E.3 Numerical Experiment 2 Nugget, (192 points) - L/40 280
E.4 Numerical Experiment 2 Nugget, (192 points) - mixed point and L/80 281
E.5 Numerical Experiment 2 Nugget, (192 points) - mixed point and L/40 282
E.6 Numerical Experiment 2 Nugget, (192 points) - mixed L/80 and L/40 283
E.7 Numerical Experiment 2 Nugget, (192 points) - point and L/80 284
E.8 Numerical Experiment 2 Nugget, (192 points) - point and L/40 285
E.9 Numerical Experiment 2 Nugget, (192 points) - point and L/80 and L/40 . . 286
E.10 Numerical Experiment 2 Nugget, (192 points) - all variograms at once 287
E.11 Numerical Experiment 3 Nugget, (192 points) - point 289
E.12 Numerical Experiment 3 Nugget, (192 points) - L/80 290
E.13 Numerical Experiment 3 Nugget, (192 points) - L/40 291
E.14 Numerical Experiment 3 Nugget, (192 points) - mixed point and L/80 292
E.15 Numerical Experiment 3 Nugget, (192 points) - mixed point and L/40 293
E.16 Numerical Experiment 3 Nugget, (192 points) - mixed L/80 and L/40 294
E.17 Numerical Experiment 3 Nugget, (192 points) - point and L/80 295
E.18 Numerical Experiment 3 Nugget, (192 points) - point and L/40 296
E.19 Numerical Experiment 3 Nugget, (192 points) - point and L/80 and L/40 . . 297
E.20 Numerical Experiment 3 Nugget, (192 points) - all variograms at once 298

Chapter 1

Introduction, Literature Background, and Objectives

1.1 Introduction

The necessity of characterizing the spatial structure of heterogeneity in soil properties and parameters is a prerequisite to many practical endeavors such as characterizing point and nonpoint source pollution, solute transport modeling, agricultural site specific management, remediation and risk assesments, and natural resource management. These natural phenomena occur in unique spatial regions during a specific time period, conditions which can never be exactly replicated elsewhere. Thus an objective analysis of these phenomena is challenging due to the inherent unique characteristics of each specific situation [Matheron, 1989]. However, from experience in studying soil physical properties under a variety of different settings, several common characteristics (i.e., to fields in general) have proved their utility to solve various practical problems. These include 1) 'relatedness' or correlation as a function of spatial or temporal separation, and 2) the dependence of the experimental histogram on sample support (length, area, or volume over which the sample is obtained). The measure of correlation, or equivalently, of variance, as a function of separation, is the most commonly used measure of the spatial structure, and is often expressed quantitatively as the variogram.

In the past, much attention was paid to point support information ignoring the effect

of support volume in the sampled data. How can we use mixed support scales to enhance variability structure estimation? The effect of collecting samples or doing measurements at a given sample volume is to reduce the observed value of the sill and increase the observed value of the integral scale, as this sample volume increases. Therefore, using measured data ignoring the effect of these facts will lead to erroneous estimates of the true point variogram. It is important then to concentrate efforts in providing tools and methods to provide a more precise estimate of the point variogram using regularized variograms obtained from field measurements, as well as estimates of the dispersion variance of the field data. The literature in the field of geostatistics is rich and has received several improvements and excellent references in recent years. For example, see the references by [Goovaerts, 1997], also [Chiles and Delfiner, 1999], as well as [Deutsch and Journel, 1997]. Many of the approaches in geostatistics are concentrated on the point variogram. The availability and affordability of computing resources has put within reach a great possibility to perform more exhaustive and rigourous studies regarding model applicability and robustness. Such is also the case for sample support analyses.

One of the first studies that used the idea of different sample support measurements to determine the spatial correlation structure of a field was [Shouse et al., 1994]. This paper employed regularized variograms and observed dispersion variances to infer the point variogram (spatial structure) of the field. This paper used real measured data from a field experiment, with two different measurement scales, one at $25cm$ by $25cm$ with 256 samples, and the second one at $1m$ by $1m$ with 16 samples. The results show the utility of the method to estimate the point variogram and so inspired further similar ideas. Follow-up work by these authors is described in [Ellsworth and Boast, 1996]. This could be considered a seminal work for the ideas worked on in this dissertation, in particular for the use of multiple sample supports, to estimate the variability structure. In addition, in this work a generalized correlation function is used that employs the mixed support variogram to calculate a type of generalized dispersion variance that identifies correlations between the same, and

different, properties measured at different support scales. This paper is probably the only one in the soils literature that shows such an approach to relate measurements made at different support scales with each other in a geostatistical way. This was a novel approach and guided the variogram fitting in this work. Further, the ideas in [Russo and Jury, 1987a] were the inspiration for the Monte Carlo numerical experiments used in the uncertainty and robustness analysis of the method.

1.2 Literature Review

Geostatistics and Spatial Variability

The use of geostatistics in hydrology and other environmental sciences started some time during the last twenty years. Perhaps the most popular book on the subject was *Mining Geostatistics* [Journel and Huijbregts, 1978]. The book was not intended for environmental purposes, it was for the purpose of mining engineering. This is a rather old book, but it was for a long time the best and most complete reference about geostatistics. It is basically a compilation of class notes from the classes taught in France, at the Fontainebleu Institute. The book is difficult to read and sometimes it takes a good amount of time to decipher the concepts behind the equations and methods presented, but the concepts are all there. There is a lot of repetition, and sometimes the material seems to be disorganized. The concepts about regularization are all well explained here, as well as more subtle and less understandable ideas such as the Cauchy algorithm to reduce multiple integrals into more manageable units. Many people think this is not a good book because it is old, repeats material, and many times it lacks clarity. But it is a great reference and was the seminal work for many years.

Later on, some more references started to appear. One of them, coauthored by Journel himself, was [Deutsch and Journel, 1997] which is now in its second edition. The book presents GSLIB, a compilation of computer programs that implement algorithms and ideas

that are presented in [Journel and Huijbregts, 1978]. The programs presented in this book have been written, tested and debugged for about twenty years, and in its second edition this book is one of the most important tools for anyone involved in the analysis of spatial variability of data. The programs alone can be downloaded for free from the Internet at Stanford University, but the book is an excellent companion, it explains the input data format as well as the underlying theoretical principles behind the programs and procedures included in GSLIB [Deutsch and Journel, 1997].

A more recent reference on a similar line of analysis but quite a bit more complete and organized is [Goovaerts, 1997]. The book is excellent, with a clear and rigorous mathematical analysis of spatial variability. The most favorite part of geostatistics for Goovaerts apparently is the use of indicator variables, and so indicator kriging, as well as sequential indicator simulation are topics clearly exposed in the book. In particular the use of hard and soft data is well explained in the book, and so are alternative approaches different from sequential simulation. Other authors have a different approach to spatial statistics. Such is the case of Cressie. His book [Cressie, 1993] has been a reference of choice in several fields involving, to some extent, the use of spatially distributed data where correlation might be an issue of interest. The book is full of mathematics and statistics, and in a way it summarizes the vast number of papers that Cressie and others have authored or coauthored during his career. The first chapters are devoted to geostatistics and so it is a good reference in that sense. It covers several topics; for instance the treatment of trends and their removal using mean/median polish is explained in good detail. Also, other fields such as remote sensing and digital image analysis are treated. This is a very large and quite comprehensive reference in spatial analysis.

Applied Geostatistics

Some other authors have included a large amount of geostatistics in references that are not directly related by themselves to spatial variability. Such is the case of [Gelhar, 1993] where

the book is titled *Stochastic Groundwater Analysis*, yet there is a chapter at the end that talks about geostatistics. This book is not long, but it is quite intensive and concentrated. Every sentence and paragraph requires a careful reading. It summarizes a large amount of information present in the literature, but put together in a clear yet challenging way. The notation used for things like summations tends to be unique and at first confusing. In any case, this is a very important reference in geostatistics.

The paper that started it all in the area of geostatistics applied to hydrology and other environmental disciplines was the one by [Delhomme, 1979]. This is a classic paper, it has been referenced several times in literature, it is fairly old but still the ideas and concepts presented in it remain absolutely up to date. This is one of the few papers that makes an appropriate presentation of the concept of dispersion variance, estimation variance, and extension variance. It also presents the concept of kriging in good detail, as well as the concept of stationarity. Conditional simulations and their uses, as well as the turning bands method and assumptions of spherical covariance functions are the key components of this work. The paper considers that, at the time, conditional simulations are fairly new and so it promotes their use and it presents them as a simulation of the Monte Carlo type, to generate 2-D and 3-D correlated random fields. The simulated field has the same correlation structure as the original data used for conditioning. The paper makes a good presentation of the issue of scale, warning that there are no point values, there are always some sort of regularization, and the concept of regionalized variable should be seen in the context of a model, an approximation to reality. Dispersion variance is explained as the dispersion of punctual samples of size x within a given domain D (which can be a subdivision of a field, or the entire field itself). The extension variance is a different concept, while dispersion variance has a physical significance, the extension variance is an operational concept: it is related to the error incurred when one tries to extend to a domain D the value measured at a given point x. The variance of this error is the extension variance of x onto D. A third variance concept, the estimation variance is not conceptually different from the extension variance. [Delhomme, 1979] makes a good

distinction between uncertainty and spatial variability: uncertainty is related to the amount of spatial variability, but it is strongly dependent on the sampling pattern. Parsimony is the use of the simplest model that describes appropriately a data set. Kriging is the estimator of choice for the linear estimation needed in this paper. The paper presents a good detail of what is first order stationarity and the intrinsic hypothesis. Also, a simple treatment of drift is included in the analysis (estimate by simple least squares), as well as the possibility of having to use a generalized covariance when the drift is more complex. The conditional simulation is presented as first a non-conditional simulation using the turning bands method, then using kriging to make estimates of the data in the more dense grid, using the initial coarse grid as the basic data set. The author uses the log of the hydraulic transmissivity because it more readily allows identifying the spatial structure than the transmissivity by itself. Two case studies are presented to show the use of the method. The paper suggests at the end not only that the transmissivity be considered in the inverse method but the head values as well. This is probably why this paper has spawned so many ideas in other papers by Kitanidis, Dagan, etc. As it seems, this paper put forward the basic ideas that these other authors picked up on: conditional simulations, trends, dispersion variances, and other ideas important in spatial variability analysis.

This is a good point to mention a book that deserves a special mention for its depth and completeness, yet it is just a few hundred pages long, [Matheron, 1989], an excellent book. This is a very short book, yet it is extremely powerful and full of information. It is extremely dense in its contents, but once there is an understanding of how to read the information it provides, then it is not so difficult to understand. The book talks about the objectivity of geostatistics. There is not a single way to explain how good the author goes about the whole issue of sampling and also about estimation, but certainly the book has given me good ideas to explore regarding types of supports and what happens to dispersion variances. At least this is one publication that makes a rigourous approach to dispersion variances. Yet the paper leaves many more doubts that one might have had to start with.

Variogram Analyses

Perhaps the first paper that sparked the idea of the work done in this dissertation is [Russo and Jury, 1987a]. It considers point variogram uncertainty only, thus the idea of regularized variograms was one of the first and most natural extensions to the work of these authors. (Also, the idea of inverting regularized variograms and dispersion variances was inspired by [Shouse et al., 1994] and [Ellsworth and Boast, 1996]). [Russo and Jury, 1987a] show an analysis with a Monte Carlo approach to estimate the width of confidence intervals in the variogram. They criticize other authors and scientists,in particular the works of Kitanidis: [Hoeksema and Kitanidis, 1985a], [Hoeksema and Kitanidis, 1985b], [Kitanidis, 1995], and [Kitanidis and Lane, 1985]. In this dissertation, the Monte Carlo approach used was not to generate equally probable realizations of a field and then doing a sampling exercise; rather, the approach in this thesis was to sample an individual field in several ways that were decided using a Monte Carlo approach. There are several points in this paper that are of great importance. For example, the definition of correlation scale by means of an integral, in 1-D and 2-D, which is also shown in [Matheron, 1989] and [Gelhar, 1993]. The paper explains clearly why the estimation of the variogram with a traditional estimator is meaningful if the distribution of the data is binormal, and presents some of the methods discussed by [Cressie, 1993] to estimate the variogram in a robust way. [Russo and Jury, 1987a] is one of the papers that explains that the simultaneous estimation of the sill and the range of a variogram will give biased estimates. This concept of bias in the simultaneous estimation of the sill and range of the variogram is further criticized in [Gelhar, 1993] where it is explained that the sill can be independently estimated with a large number of samples separated at a larger distance than the range, as well as an examination of methods to estimate the variances of the sill and range estimates themselves. However, uncertainty in the sill creates uncertainty in the range, which is exactly the concern about bias if the range and sill are estimated simultaneously. In any case, [Russo and Jury, 1987a] chose to use the

Maximum Likelihood (ML) and Restricted Maximum Likelihood (RML) methods of parameter estimation, cited elsewhere in literature (for example, [Hoeksema and Kitanidis, 1985a]) The authors selected the variogram model to use based on the Akaike Information Criterion (AIC) which is mentioned in several parts in the literature. To test for the normality of the data, the authors used the Kolmogorov-Smirnov test (KS) which is referenced in several references in the literature. One of the most important concerns in [Russo and Jury, 1987a] is the fact of biasedness in the simultaneous estimation of sill and variance. They conclude that bias has greater or lesser importance depending on the purpose of the estimation; if the variogram is going to be used in estimating conditional probabilities, then it is recommended to do an RML estimation of $b = Sill/range$. But if the purpose of the estimation is to estimate unconditional probabilities (such as stochastic transport modeling), then the values of the sill and the correlation range are needed individually, which might pose serious problems due to their uncertainty. On the other hand, if the need is the underlying theoretical model or ensemble, then one set of field samples is not very enough. One last note about this paper has to do with accuracy: the accuracy of the estimate of the correlation length depends on the ratio of the underlying integral scale to the size of the sampling grid Δ and on the distribution of the separation distances, or more precisely, on the number of pairs separated by distances smaller than the range R.

The second part of the series of these two papers is devoted to the case of non-stationary fields. In the paper, [Russo and Jury, 1987b], the authors do a similar analysis to that of the first part, but in this case they include drift. They start by mentioning what kind of known drift is easy to remove and what methods are there available that allow the removal of drift. The methods available are: Generalized Covariance (GC), Iterative Generalized Least Squares (IGLS), and the RML mentioned quite a few times by Kitanidis (for example, [Hoeksema and Kitanidis, 1985a], [Kitanidis and Vomvoris, 1983]). All the methods, as presented by Kitanidis, assume that the drift can be represented by a local scale or within a local neighborhood as a linear combination of linear polynomial functions, in a way that

the removal of the drift is specifically accomplished as part of the algorithm. The RML and GC methods are similar in how they treat the drift component, making the IGLS method stand out as an alternative for drift removal. This method is a trend analysis method. In simple words, different polynomial degrees are tried out and then the next step to build upon the current polynomial. The authors basically do a study of drift with different assumed types of drift functions. They consider linear and non-linear drift, and do some variogram analysis using the Fractal Dimension concept. The estimates are actually obtained with IGLS, RML and ML (maximum likelihood), which is essentially everything that is available in the literature. The validation methods were several: MSD, OLS (ordinary least squares) and MSD, IGLS and MSD, ML and RML. The criterion to select a model was the same as in [Russo and Jury, 1987a], the AIC (Akaike Information Criterion). This is an interesting paper, but the authors assumed a particular drift function and so the paper does not present a general methodology. There are other ways to remove the drift, but the only method presented here is in the variogram fitting procedure.

Regularized Variograms

It is important to realize that the idea of using the concept of $\bar{\gamma}$ has been pointed out in the literature several times for the purposes of regularized data and change of scales. The reference by [Clifton and Neuman, 1982] uses regularization that is just averaging over an area. Pump test data sets are used, disregarding (but noting) that leakage or partial penetration are possible sources of error. The use of $\bar{\gamma}$ is extensive and in detail for purposes of showing how to transform values of hydraulic conductivity computed with the pump test data and the corresponding effects. Some mention to dispersion variance concepts is presented too. A similar paper, but at a much higher level of mathematics is the one in [Desbarats, 1994], it is clear, to the point, and it contains a good analysis of transforming measurements of hydraulic conductivity by means of pump tests into some more meaningful results because of the implicit regularization resulting from a pump test is done (i.e., the value that we are

measuring is actually an area/volume averaged one). It makes use of $\bar{\gamma}$ and shows a clear and real applicability of the issue of regularization. 3-D radial flow to a pumping well is analyzed, and essentially gives a method to scale hydraulic conductivities measured at the borehole to the field scale. The method is tested using a numerical model showing excellent agreement. The second moments of the "drainage scale hydraulic conductivity" are computed. The simulated fields were created using the turning bands algorithm [Journel and Huijbregts, 1978], with an exponential model of correlation and isotropy. The method is clear but cumbersome to use. Another example of the use of $\bar{\gamma}$ is presented in [Sánchez-Vila et al., 1995] where a comparison among three methods for upscaling hydraulic conductivity in a non-linear fashion are compared.

The work of [Zhang et al., 1990] considers block variograms, definitions, dispersion variance values, and analysis, but the authors take another approach using auxiliary functions for their rather accurate computations. Methods are presented to change support (using numerical integration), using spherical and linear covariances. The paper analyzes and evaluates the effect of sample support size on variance. The results are compared to those of the classical relationship developed long ago (1938): the variance is reduced from V to V/n^b as the support area increases from 1 to n plots for uniformity trials. The exponent b varies from zero to one. The results are given in terms of approximations by rational functions for ease of calculation. Part of the applications include uniformity trials and also measurement theory. The formula presented above is equivalent to: $V_W = V_w \cdot [w/W]^b$ if n equals the ratio of area W to the unit support size w, where V_W and V_w correspond to the variances of the two areas. The situation $b = 0$ corresponds to the case of $V_W = V_w$, and $b = 1$ to the case of completely random sampling. Only isotropic models are considered, and in some cases a second order stationarity assumption is considered (which implies the existence of a finite population variance σ^2). Most of the cases use a spherical model, but an exponential model is also used to present the concept of range vs. correlation range (or effective range

as it is sometimes called). The paper has an excellent table showing common variogram parameters, with their name, equation type and the effective range. Some unusual and not so common variogram models are presented in this reference, for example the Michaelis-Menton model, whose function is: $\gamma\left(\frac{h}{a}\right) = \left(\frac{h}{a}\right) / \left[1 + \left(\frac{h}{a}\right)\right]$. The effective range is at h/a $a = 19.0$. All the effective ranges computed were based on the value of h where 95% of the sill is attained. The models studied here include a nugget. The variance decreases as the sample support increases, which is consistent with the results elsewhere in the literature. Both the sill and the nugget decrease as sample support increases. The authors clearly explain that the regularized variables (or averaged variables) have an inter-relation with variances at other scales, and the relationship is given in terms of the dispersion variance. The authors introduce in a clear and simple way the concept of Krige's relation, which relate multiple scales. Several approximations and simplifications of the basic formulas for both dispersion variance and regularized variograms are presented, and so the final approach in the paper is to combine the approximation with numerical integrals so that the final result is to obtain more efficiently regularized variograms. And as a final note, the point variogram is estimated by transforming the sill and the range but not doing any numerical optimization process. Therefore, in general, the relationship between the point variogram, regularized variograms, and dispersion variances is explicit and direct but complex, several uses and applications are available in the literature but the most important point is that it allows for the development of a scheme to use, simultaneously, different types of information to fit point variograms and other applications. [Ellsworth and Boast, 1996] computed the experimental variograms using the traditional estimator presented in Chapter 2 for the mixed support and regularized variograms. The computation of the experimental dispersion variance was very interesting and straightforward. Basically, the authors used the cores data, as well as the composite (donuts + cores) to subdivide the field into all possible arrangements of equally sized blocks subdivisions. The next section formalizes the optimization scheme for the point variogram parameter estimation formulation.

Regularized Variogram Deconvolution

The methods presented as part of this dissertation regarding point variogram fitting using regularized variograms (single and mixed sample support) as well as dispersion variances, have been published in [Warrick et al., 1999]. The purpose of that paper was not to present the fitting method alone, which was already presented in [Ellsworth and Boast, 1996], but to present an insight into spatial variability assessments, including (but with other ideas too) the issue of sample support size effect. [Warrick et al., 1999] presents the idea of estimating the point variogram from regularized variograms and also using dispersion variances. It presents some considerations about non-point source pollution, sampling strategies, population domains, alternative experimental designs, some deconvolution of experimental regularized variograms, two dimensional upscaling and downscaling. The paper discusses in some detail the issue of conflict of objectives: we want reliable measures using limited resources available. Practices such as precision farming (site specific management) require a detailed spatial variability assesment of the farming site. The purpose in this case is to help the precision farmer in reducing the amount of chemicals or nutrients required in the field and thus reducing costs and environmental damage. Environmental damage reduction can be seen from the reduced usage of chemicals so the soil and groundwater in the area will not receive so much percolation as with a traditional farm management practice.

Soil heterogeneity is of great importance in the issue of spatial structure estimation because sample support is an issue that will affect the estimates of the variograms. Models are needed to describe processes over a range of scales, and the input parameters as well as model predictions need to be estimated for the appropriate scale.

Field Observations

The first example presented in more detail in Chapter 4 in this dissertation uses some of the data presented in [Warrick et al., 1999]. The experiment site which is referenced several times in this thesis, is in [Ellsworth and Boast, 1996]. It is a $4mx4m$ plot of Pachappa fine

sandy loam at Riverside, California. Pulses of chloride, nitrate and bromide were added successively using a special uniform sprinkler at a rate of $1cm/d$ net applied water (NAW). The solute pulse corresponded to $0.29cm$ of NAW for chloride, $3.29cm$ NAW for nitrate and $4.78cm$ NAW for bromide. The total NAW at time of sampling was $27.5cm$ for chloride, $20.1cm$ for nitrate and $12.5cm$ for bromide. The center of $2mx2m$ of the plot was sampled using destructive sampling at the center $2mx2m$ of the plot. Layers of $10cm$ increments were sampled to a $2m$ depth. Each layer consisted of $9x9$ cells of equal size, each of area $22.2cmx22.2cm$. The cell was composed of a core of $7.4cmx7.4cm$ centered, and a donut. So the result were 1620 cores and 1620 donuts (each sampled independently). Experimental variograms were computed based on the total chloride recovered within each of the 81, $2m$ long vertical columns at each measurement support. The experimental variograms are regularized, thus they underestimate the point variance of the soil property within the field. It is possible to estimate the point variograms from the regularized ones and then get an estimate of the true variability structure (assuming second order stationarity) as done in [Ellsworth and Boast, 1996]. The equations to perform the deconvolution process are presented in Chapter 2 in this dissertation. The use of Krige's relation is also noted and is the same formula stated in Appendix A of this dissertation. The variogram model used is the exponential model, and the method to do the data fitting is a generalized least-squares algorithm (Levenberg-Marquardt). Maximum concentration levels are affected, as discussed in [Ellsworth and Boast, 1996], by the sample support size: when the support increases, the maximum measured concentration level will in general decrease. One thing to note is that inferring the shape of the histogram for samples at an arbitrary support V from measurements at a smaller support v can be achieved by stochastic simulation of the v scale process. Using measurements at multiple supports, as discussed in [Ellsworth and Boast, 1996], provides a robust method for estimating the spatial structure of mass recovery. Samples at the smaller support are more sensitive indicators of the true or point support spatial variability, while larger supports are more efficient in estimating field-average properties.

[Warrick et al., 1999] presents another example regarding the Williams Field, which is presented and discussed in detail in section 4.2.2 of this document. At this point, only a few details need be noted: there was a 250 $\frac{pound}{acre}$ cutoff laboratory limit for phosphorous measurements (thus leading to the decision of not using any data value which was recorded at 250). Also, many things are said in the paper about sampling: first, that sampling at larger support typically provides a more robust estimate of field-size trends. But it does not provide a direct estimate of the point variogram. Second, it is thus possible to estimate the point support variogram from a combination of two distinct larger support variograms. Therefore sampling at large supports is a way to obtain a robust estimate of field size variability, as is also discussed in [Ellsworth and Boast, 1996]. This combined with the method of inferring point support variograms, make up a robust and cost effective approach to estimate small scale variability as well as field scale trends at least for separations larger than the support used. In reference to sampling patterns and their influence on variances, the authors used a $1024x1024$ simulated field and four different basic sampling patterns to study the experimental field-average properties and dispersion variances. Determining the best sampling strategy was not the main purpose of the study, rather it was to give some results on what to expect using several experimental sampling strategies. The sampling strategies were:

1. Quasi-point support area in the center of each grid cell.
2. Sampling along the diagonals of the grid cell.
3. Diamond shaped sampling.
4. Block average sampling for each grid cell.

Estimates of the field average become increasingly uncertain if fewer points are sampled. The block average remains the same for all cell sizes since it is exhaustive (uses all the information available). As an interesting note, sampling strategies 2 and 3 somewhat diverge from the

true field average if the field is divided into $4x4$ or less grid cells. The reason behind this is not really the effective sample size, rather it is related to the sample shape and to the existence of a spatial structure in the field. The dispersion variance, $D^2(v/V)$, of grid cell support becomes difficult to estimate for unusual patterns (v is within field V). It is defined by the authors as the between blocks variance in the overall field. The minimal requirement for estimating the dispersion variance is one sample per grid cell. The use of diagonals or quadrilaterals provides good results for dispersion variances. A second numerical simulation was done, where the exact same set of analysis were performed. The result was that the dispersion variance decreases more rapidly as the cell size increases compared to the original case (original range=$210m$, new range was $105m$). The effect of sample support on "Threshold Values" is well known in mining applications, where certain material like ore is mined if it is expected to exceed a cutoff value. In a similar way, nutrient application in agricultural applications can be regarded as a cutoff value problem: certain nutrients such as phosphorous will be applied if the measured level is less than 40 lbs/ac. The area of land below the threshold will depend on the area of the supports. For a traditional management operation, the entire field is treated equally and so only the entire field threshold would be of interest. However, in precision farming management, the unit chosen as the management unit will dictate the nutrient application (fertilization) strategy. Thus the total area to be fertilized and the amount of fertilizer needed will be dependent on the management unit to use. As the sample support size becomes larger, variability is reduced (the averaging involved in selecting a larger volume support smooths out the natural variability of the field). If regularized variograms are available, then the point variogram can be estimated. This also holds for when dispersion variances (experimentally computed) are available. In any case, with the availability of the point variogram, then dispersion variances can be computed for any kind of sample support.

Variogram Uncertainty

The whole issue of uncertainty and sampling was the next step in the work for this dissertation. Once the method was developed, there was the issue of robustness and uncertainty. In other words, the question to answer was then: how good is the methodology? to do this, a similar path was followed from other references in the literature, in particular, the work presented in [Shafer and Varljen, 1990]. The authors do an analysis of computing confidence intervals of the semivariograms using both a jackknife estimate and a comparison/verification it using a Monte Carlo approach. The authors did some analysis of regularized variograms, but they did obtain the point variogram from the block variograms. The analysis done was related to sources of uncertainty by the use of regularized variables. The authors chose the jackknife estimation to avoid or reduce the bias in the estimation of the variogram parameters. The authors use the traditional sample variogram estimator $\gamma(h) = (1/N(h)) \left\{ \sum_{i=1}^{N(h)} [z(x_i + h) - z(x_i)]^2 \right\}$ This paper is one of the few papers that make a clear distinction between cross validation and jackknifing: Cross validation (which is sometimes mistakenly referred to as jackknifing), consists of leaving one data point out and checking how well the rest of the points estimate the point which was left out. The jackknife estimator used in the paper is basically a technique where the entire data set is partitioned and then the variability is calculated by the weighted differences between the parameter estimated with the entire data set and the parameter estimated from the entire data set excluding one of the partitions of the data set.

Scale Issues

One of the important issues in this dissertation has to do with change of scale and sampling designs. The literature is rich in this respect, several investigators in environmental fields have at least once mentioned issues of scale, change of scale and upscaling. It is not an easy task to pick a subset of the vast amount of literature covering the issue of change of scale. The paper by [Addiscott, 1993] deals with several change of scale issues, in particular

the use of a certain scale for model calibration and then the use of the same "calibrated" model for estimation and simulation purposes, disregarding the implicit/explicit change of scale issues. The paper summarizes a fairly recent validation scheme, and also suggests the fact that models should be analyzed in terms of the sensitivity to not only changes in their parameters but to changes in the variance of their parameters as well, in particular when the models are non-linear with respect to their parameters. Non-linearity issues are of special importance when the parameters are subject to some spatial averaging or interpolation. The whole issue of spatial dependence of parameters, as would happen in a model that uses soil information, is enough reason to consider the variation and the mean of a parameter for sensitivity purposes, not only the mean, as is usually the case. This paper has a fairly long discussion about model classification issues, giving a short but useful table of model classifiction based on whether a model is deterministic or stochastic, and a subclassification for each, whether a deterministic model is Mechanistic or Functional and whether a stochastic model is Mechanistic and Non-Mechanistic. Other considerations for model classification include complexity, flexibility, and transferability.

A more sophisticated approach to the issue of change of scale is followed by [Neuman, 1990]. The problem dealt with in the paper is that laboratory scale analysis of a field scale phenomenon is an incorrect approach. The author proposes a more general method that uses a fractal dimension. The method proves to be quite general and difficult to apply. The variables considered must heavily are the dispersivities and the fact that they increase with scale. It gives a good insight of four different scales that concern groundwater literature: pore, laboratory, local field, and regional scales. The paper shows that a fractal model of log hydraulic conductivities is consistent with the scale effect exhibited by field and laboratory dispersivities. The scaling rule, which in a mean sense applies universally over the broad class of geologic media and the wide range of scales represented by the available data, is mathematically correct but not of much practical use. The mathematical analysis starts by proposing an anisotropic exponential covariance model. The mathematical analysis of

the authors leads to the conclusion that the field data represent a self-similar log hydraulic conductivity field, and that the variogram model is approximately $\gamma(s) \cong cs^{0.5}$, where c is a constant and s is distance. This semivariogram is characteristic of a random fractal with dimension $D \cong E+0.75$, where E is the topological dimension of interest ($E = 1$ for a linear transect, $E = 2$ for a 2-D plane or cross section, $E = 3$ for a 3-D medium). The relationship between D and $\gamma(s)$ is not obvious; the relationship can be explained first by saying that in the general case $D = E + 1 - \omega$, where ω is bounded between 0 and 1. The quantity ω is called the Hurst coefficient. When $\omega = 0.5$, the property is not correlated in the E dimensional space, resembling Brownian motion. When $\omega > 0.5$ and $D < E + 0.5$, then the property is positively correlated.

The work in [Beckie, 1996] is probably one of the best analyses of sampling, resolution and bandwith available from a set of available measurements. This reference offers approaches on how to accomodate a change of scale by means of mathematical filters, there are some geostatistical examples and relationships of cokriging and Kalman Filters. The idea of expressing the sampling process as a spatial filter allows use of spectral techniques to assess the aliasing and resolution associated with the sampling method. Suppose there is a grid of measurements, and if secondary values are available to compute a continuum of it (such as a map) then cokriging can be used. For example, slug tests and pump test conductivities would be each represented as spatial averages of a core-scale conductivity (i.e., one-dimensional regularization). The filter functions can be used in geostatistical estimation procedures to condition estimates of one parameter with measurements of another. The energy spectrum (Fourier transform) requires a dense, regular sampling to estimate this, and essentially displays the scales at which a function varies. The measurement scale limits the information on how the parameter varies on subgrid scales. Model parameters are smooth functions because they represent the effects of spatially averaged physics. When the full detail of the parameter field is desired, the influence of the maximum sampling interval on the parameter field resolution is the subject of results known as *sampling theorems.*

These theorems say that smooth functions with a bandwidth $f_b = 1/2\Delta_N$ can be exactly reconstructed from discrete samples taken at locations that are separated by the Nyquist interval Δ_N. When the measurement support is fixed, then sampling at intervals smaller than the Nyquist interval Δ_N provides no new information, but can be useful if the data are noisy. Sampling theorems also show that band limited functions cannot vary arbitrarily below a certain length scale and remain smooth. Therefore, they can be smoothly interpolated without error between sample points that are separated by less than the Nyquist interval. Similar theories address nonuniform sampling grids. For band-limited parameter fields the appropriate Nyquist interval equals the measurement scale λ_M. In conclusion, if one wanted to measure a site in maximum possible detail with a maximum number of tests, then the samples should be separated by distances of the order of λ_M [Beckie, 1996].

The whole theory of using filters can be cumbersome at times and obscure to say the least. There are a good number of references in the literature about more specific and applied research using filters and that incorporate a wide variety of already available measurements. The good effect of such practice can be seen in [Petach et al., 1991] where the data available and used by the authors included soil surveys, remote sensing data, land-use databases, and pesticide handbooks, in the analysis of the variability of the estimated leaching potential of some agricultural chemicals. The way used to incorporate the previous data is by means of pedotransfer functions so that all the data are transformed into a more homogeneous dataset. The model LEACHM (Leaching Estimation And CHemistry Model) is a one-dimensional finite-difference model designed to simulate the movement of water and solutes through both layered and non-layered soil profiles. The time step is variable, and is calculated based upon allowable water content changes in the profile. It has been used in many applications, both for prediction and validation. The model, in the version used in [Petach et al., 1991], did not take into account any preferential flows (macropores, worm- holes, other structural formations). The pedotransfer functions used took into account percentages of clay, sand, and silt to describe the values of hydraulic properties.

Perhaps the most complex and complete reference on the issue of small and large scale effects and nesting is presented by [Cushman, 1986]. It presents the relationships among different scales of measurements, laboratory scale, and measurement device scales. Geostatistics are discussed in it, and it treats the entire issue of measurement device scale as a convolution problem. The author criticizes the traditional approaches of neglecting or inappropriately considering the scale issues, as well as designing measuring devices that do not really fit the needs of the measurments to be taken in the field. This paper also emphasizes the fact that regularization erases the high frequency/lower scales information (nugget effect in particular) and so makes it difficult or impossible to accurately recover such information. According to the paper, it is not a good idea to mix measurements from different measuring devices or methods unless there is good accounting of the fundamental scale of a given device (the upper frequency response bound). The fundamental scale of the device (or method) is what really determines the minimum realistic distance between measurements that would make sense in the field. The uncertainty created by these measurement gaps can be incorporated into a model by either of two ways: using a mathematical filter procedure or by conditioning the desired results in a probabilistic way to existing data alone. The way [Cushman, 1986] defines device really is a mathematical definition, more of a method itself rather than a physical measurement device. The whole mathematical development is based on the idea of convolution, which goes back to systems theory where a signal has to travel through a medium or a series of components, thus the response is the convolution of all the effects of the components. There is a quite detailed presentation of linear interpolators as filters, and the linear filter of choice is that of kriging. But at the end, the paper becomes critical of ideas used in this dissertation. Basically, [Cushman, 1986] does some mathematical analysis and explains that the variogram cannot really be estimated at the origin and the reason is that a) It makes no sense to measure quantities at scales smaller than the scale of the apparatus, b) the minimum lag distance is in general larger than the instrument's fundamental scale. However, the point is that there is a large degree of uncertainty

when extrapolating the regularized variogram results to compute the point variogram. In particular, there is a problem when the nugget effect is regularized, because what happens is that the regularized nugget goes to zero almost instantly with any sample support size [Goovaerts, 1997]. As was seen, the issue of nugget effect is affected by regularization more than large scale trends. In fact, several authors and references treat the issue of large scale trend representation and removal. Many methods exist to accomplish these purposes, from something as simple as to consider the trend to be simply a nested structure with a very large range, to more sophisticated techniques assuming a certain function that is unknown, or of a global polynomial form.

[Heuvelink, 1998] argues that in general the so called pragmatic approach to treat the problem of change of scales is preferable to the complications and mathematical burden of developing the inter-scales relationships for the specific situation. By pragmatic is meant that the data availability, support and dominant processes are the starting points for the solution of the problem. It should be clear that this approach is too specific for each scale and most likely a different model would result for each scale. This does not mean the procedure is invalid, it can just be less efficient but effective. But sometimes models that are supposed capable of being adjusted for different scale have problems because the people who specified such changes and modifications inadvertedly forgot to consider certain issues. In such cases, an overly complicated model can result because data availability was not considered correctly. A good tool is avaialble to assess whether or not simplifications were correctly made, error propagation analysis, also called uncertainty analysis. This tool not only checks but also serves a purpose in giving insight about what direction to modify the model. The aim of error propagation analysis is to determine how large the error in the output is, given the errors in the inputs. Ideally, error propagation analysis should also include model error. It can be included by either making model parameters randomly distributed or by adding a noise term to the model.

Trend Removal

A good reference about trend removal is [Rajaram and McLaughlin, 1990]. The paper presents an algorithm in two versions, one is the discrete version where the data are limited and probably irregularly spaced, and the other one is the continuous version which is probably of less interest but can be useful for the analysis of the effects of using incorrect spectral parameters. The discrete algorithm is used with more detail, analyzing both real data and synthetic data, yielding satisfactory trend estimates. The authors show that the estimation error lower bound depends on the scale disparity (i.e., $\frac{\text{trend correlation scale}}{\text{residual correlation scale}}$) and the signal-to-noise ratio ($\frac{\text{trend variance}}{\text{residual variance}}$). A good point is that these ratios may be used to evaluate the feasibility of trend estimation and removal before the field samples are actually collected. It is good to note that the discrete version of the method presented is essentially a distributed parameter Kalman filter, and it is interesting because it considers the issue of regularized samples, which is of great importance in applications. The paper relates geostatistics, Kalman filters and least squares techniques in the description. The continuous version of the method is actually a Noncasual Weiner Filter, and it is of not much practical application because the basic assumption behind it is that the property being considered is a continuous variable. The discrete and continuous methods have a very similar behavior when the number of points becomes very large, which is reasonable. Normally the trend will be assumed to be some sort of an AR1 (autoregressive model of order one in time series), which is more of a convenience. The practice of assuming an AR1 type of model for either errors or trends is rather common in the literature because the covariance function is an exponential model so its analysis becomes more straightforward.

Other authors have made a different assumption about trends. [Russo and Jury, 1987b] assumes a polynomial type of function that represents a global trend of the data. One such paper that is more relevant to applications work is [Hoeksema and Kitanidis, 1985a] and also [Hoeksema and Kitanidis, 1985b]. These papers treat the spatial variability structure estimation of 31 aquifers in the horizontal plane, for each of them the nugget, variance,

sill and range are estimated using an exponential covariance function. The authors do this analysis for transmissivity, hydraulic conductivity and storage coefficient, with the variables as they are and also with their logarithmic transformations. The paper is basically an application of the methods that Kitanidis has proposed elsewhere in the literature, where he uses his own method to remove drift (assuming the drift is a linear polynomial function). The nugget variance here is represented by a Kroenecker Delta function, like this: $\gamma = C_n \cdot \delta(h) + C_o \cdot exp\left\{-\frac{h}{l}\right\}$. The term δ_{ij} is the Kroenecker delta and is zero whenever h is not zero. The paper presents three different models: 1) No drift, log transformed data; 2) Linear drift, log transformed data; and 3) Linear drift, no data transformation. By removing the drift (large scale trend) before estimating the variogram model parameters, the bias is reduced because there is no need of estimating the drift itself. However, [Gotway and Hergert, 1997] points out that any trend removal, followed by a variogram estimation, results in a bias in variogram estimates. This is avoided with the generalized covariance, which does not explicitly remove trend. The work in this dissertation did not include any trend analysis and removal, but it is a good point to keep in mind. [Hoeksema and Kitanidis, 1985b] is an application paper, so the authors devote some time to explain higher moments and their estimation from sampled data. The data sets come from everywhere in the United States, several from Illinois, but with a strong favoring of the western part of the U.S. As a final note, it is worth mentioning other ideas for large scale trends treatment. The work in [Cressie, 1993] shows the use of the Mean Polish and Mediam Polish to remove the drift from a regular grid of data. The procedure does not make any assumptions regarding the form of the drift function, but tends to remove more than the drift itself. In any case, it is a possible method and has been used by others.

Nested Structures

One of the issues when dealing with data gathered from real fields is that of several spatial structures at different scales. Large scale trends are of particular interest because they need

to be removed before any conclusions can be made. An excellent reference on this issue is the one by [Rajaram and McLaughlin, 1990] It deals with trends, their identification and removal from data sets. The paper provides a review of techniques and theoretical basis for trend identification and removal. The method that the authors propose does not make any assumptions at all about the form of the function that represents the trend, it is a method that uses the spectrum (using a Fourier transform [Press et al., 1992]). But the problem with this approach is that a very dense sampling on a regular grid is needed to get a useful measure of the spectrum, so sampling requirements become expensive to be able to use this. This paper is again one of the few papers, together with [Chiles and Delfiner, 1999] and [Cressie, 1993], that states that variability and scale at different levels is a subjective matter. For example, at a very small scale, the trends that we see can be viewed as a nugget effect at a larger scale. The paper assumes that both the trend and the residual are stationary random functions. Regarding trend removal, a number of studies in the literature include geostatistical considerations in the inverse problem in groundwater flow. The work by Peter Kitanidis has devoted a large amount of effort to this kind of approach. His methodology considers trend removal under the assumption that the trend is a linear combination of several polynomial functions. This is unfortunately not a very useful technique because it is a local trend removal methodology, and so it does not remove the global trend. Following is a sample of his work:

- [Hoeksema and Kitanidis, 1985b] In this paper, the authors present the basics of the geostatistical approach to the inverse problem in groundwater, and then present the previous work in stochastic partial differential equations. The method used is essentially perturbation analysis, which has been applied by [Dagan, 1982], [Dagan, 1985], and [Gelhar, 1993]. The idea is to consider separately the variables involved, in this case piezometric head and transmissivity. At the end, an equation relating the covariances of the transmissivity and the piezometric head can be developed.

- [Kitanidis and Lane, 1985] This reference uses a maximum likelihood (ML) method with the Gauss-Newton algorithm for parameter estimation. It presents an idea to remove unknown drift from measurements, and it presents the Restricted Maximum Likelihood (RML) which should give estimates of the parameters that are invariant to the drift coefficients. It uses the exponential covariance model for geostatistical spatial modeling. This is perhaps the paper by Kitanidis that best describes the advantages of the ML method: asymptotically, the estimates are unbiased, minimum-variance and normally distributed. One of the main points in the paper is to present the details of the implementation of the Gauss-Newton method in ML estimation of the covariance function coefficients as well as drift coefficients. The method is very well suited for the estimation of covariance function parameters which are not affected by unknown drift coefficients. For the application of the ML method, the joint probability distribution of the parameters to be estimated has to be known. The most common assumption for this is the normal distribution. However, the ML estimation is equivalent to non-linear estimation which is a hard problem to solve. If there is a high degree of correlation among the parameters being estimated, then this approximation yields a matrix that is close to singular and so the algorithm needs to be improved. The paper devotes several pages to the details of the implementation of the ML estimation with the Gauss-Newton algorithm. The paper then goes into more detail of the ML estimation applied to simultaneous drift and covariance function parameters. The method provides, by the way, an algorithm to remove the drift from the data, but under the assumption of a linear polynomial function within each local neighborhood to represent the drift. The authors discuss that the plain ML estimation procedure underestimates the variance and integral scales if used in a simultaneous estimation of the drift coefficients, but that if Restricted Maximum Likelihood is used, then the estimates are practically unbiased but with a higher variance. Finally, the authors show results from some numerical experiments using an exponential covariance function and compare it to

other methods.

- [Kitanidis, 1995] This paper is a generalization of the previous papers by Kitanidis and other authors. He proposes a Quasi-Linearization technique which is useful in the groundwater inverse problem in cases of high variability (using always the log of the hydraulic conductivity) but with enough measurements so that the variance of the estimation error is small. The paper improves the linear approach to the inverse problem presented in a much earlier paper [Kitanidis and Vomvoris, 1983] which mixes geostatistical concepts with the inverse problem in groundwater. It is a good paper, the authors simplify the method by not including the drift in the problem. The estimation methods used are the Maximum Likelihood (ML) and the Restricted Maximum Likelihood (RML). At the end, the paper gives a very simple example in 1-D, using the exponential covariance funtion as the underlying spatial structure model.

Non-linear upscaling

Certain variables and phenomena in environmental sciences do not upscale correctly by a simple integration or averaging process. Such is the case of hydraulic conductivities, as can be seen in [Sánchez-Vila et al., 1995]. This is the only paper that deals with the problem of upscaling hydraulic conductivities (so that the regularization formula makes it consistent with energy balances). The paper basically presents several methodologies available to do this. Works by Dagan have presented a stochastic methodology for heterogeneous anisotropic media by means of the dissipation energy function, but they provide no practical method for actual computations. However, the approach by Dagan states carefully a set of conditions that most of the alternative methodologies comply with, and also that the works of others, such as Gómez-Hernández and Rubin and Desbarats, are consistent with the general requirement of the correct equivalent conductivity for the point values as well as the block values. [Sánchez-Vila et al., 1995] propose an alternative approach to computing the block conductivities, in which they state very clearly that the average behavior at a given scale is

not reproduced by arithmetically averaging local values of aquifer parameters. This paper makes a good introduction to the whole idea, background and need of upscaling parameters such as hydraulic conductivity. We need them because we are interested in processes that occur at a scale larger than what we measure, or, even more, we want to use a model whose data requirements are blocks of a certain size, which possibly is larger than the measurement support available. Thus, we need to upscale, keeping as much as possible the global behavior of the aquifer while keeping the local behavior as close as possible to reality.

On a similar line of work, the paper by [Gómez-Hernández and Gorelick, 1989] deals with the problem of upscaling hydraulic conductivity. This paper has been referenced several times in the literature on the inverse groundwater problem, as well as in the issue of scaling hydraulic conductivity, finding a value for the effective hydraulic conductivity, etc. It uses variogram analysis, assumes an exponential covariance model, makes an assumption of second order stationarity, and works with a Monte Carlo type of approach with nine examples of conditional and unconditional simulations of correlated random fields using the turning bands method. It has a good discussion and conclusions on how to select a value of the effective hydraulic conductivity. The Monte Carlo approach used by the authors is well justified because of its simplicity of use, even though it can be quite problematic in terms of computational resources. Earlier works in the literature had not considered spatial correlation, but the authors in this work do include that property between block values. The paper goes through quite a bit of explanation that the case of hydraulic conductivity is not easy to analyze, and that the geometric and harmonic mean of the hydraulic conductivity (which by the way is assumed to be log normally distributed) in the case of 2-D aquifers are extreme cases of the actual value.

In an unsteady flow case, [Gómez-Hernández and Gorelick, 1989] show that the hydraulic conductivity depends on time. Early works using Monte Carlo simulation proposed a power averaging method of scaling the hydraulic conductivity, using an indicator variogram. Since one of the earliest and easiest approaches to analyze pump test data is using the Theis

equation, [Gómez-Hernández and Gorelick, 1989] reference works where the effect of spatial variability using the Theis equation is studied. A general conclusion of that approach is that angular variations have a large effect on the results. The conclusions from the nine examples are that the effective hydraulic conductivity can be taken as the geometric mean for the 2-D case, above the geometric mean for 3-D and below for 1-D. The harmonic mean is the lower limit to the estimates of the effective hydraulic conductivity, K_{eff}, and the upper limit is the geometric mean. The resulting value, is then some value in between the harmonic and the geometric mean. The following equation is proposed (which is indeed what the literature has used, in particular [Sánchez-Vila et al., 1995]): $K_{eff} = \left[\frac{1}{N}\sum_{i=1}^{N} K_i^p\right]^{1/p}$ The value of p ranges from -1 (harmonic mean) to 1 (arithmetic mean). The expression is undefined for $p = 0$, but the limit as p approaches zero is the geometric mean. Another Monte Carlo analysis was performed to obtain an approximation to the value of p (with 20,000 simulations) and the value found is indeed between the harmonic and the geometric mean. The key factor is the coefficient of variation. The final conclusion seems to be that the geometric mean is a good approximation if we are away from the wells. The geometric mean does not work near the wells; around the wells, a value between the geometric and the harmonic mean should work better. And, finally, there is proof that there is no effective hydraulic conductivity value for two-dimensional radial flow based only on the spatial statistics of hydraulic conductivity.

Sampling

An important aspect in the computation of the sample variogram is the possible selection of locations to reduce the variability of the estimates of $\gamma(h)$ when a constraint is imposed on the possible number of locations to use. There are several approaches to this problem, and the works of [Russo, 1984] and of [Warrick and Myers, 1987] are briefly mentioned here. Both methods are closely similar with few exceptions. The aim of the project in [Russo, 1984] is that the sample variogram is usually very irregular, due to natural anisotropies, unequal number of pairs per lag and the variability of the values of the squared differences of the

observed values within the averaged lag class, which increases with lag separation. To solve these problems, the author suggests a different approach to the computation of the sample variogram as a network design problem, where the issue is now how best to divide the field into a sufficient number of lag classes such that each contains an equal number of pairs of sampled points, and at the same time keeping the variability of the values of the squared differences of the observed values within the lag class as small as possible. All of the restrictions just mentioned can be incorporated in an optimization problem where the uniformity of the values within a class is measured in terms of a factor which is similar to the dispersion variance. The work done in [Warrick and Myers, 1987], on the other hand, is a more practical approach. For example, the paper says that the number of pairs for both short and large lag distances is generally small, and so that the largest number of couples occur at approximately half the maximum separation distance. The method proposed then is to select the sampling sites, in a multi-objective optimization kind of fashion. Ideally, one desires that the following list of objectives be all met simultaneously, but of course in reality these objectives are in conflict:

1. For each distance (and angle for the anisotropic case) class, the number of pairs should be as large as possible, particularly for short distances.
2. The average of the distances in each class should be close to the plotted lag.
3. The variance of the distances in each class should be small.
4. The average of the angles in each class should be close to the plotted angle.
5. The variance of the angles in each class should be small.

It is good to note that a regular grid would tend to ensure conditions 2 to 5, but only a large number of points in the grid would really help ensure condition 1. The formula to be minimized is:

$$SS = a \cdot \sum_{i=1}^{NC} w_i (f_i^* - f_i)^2 + b \cdot \sum_{i=1}^{NC} m_{1i} + c \cdot \sum_{i=1}^{NC} m_{2i} \tag{1.1}$$

Where: $w_1, w_2, ..., w_{NC}, a, b, c$ are user selected weighting coefficients, m_{1i} are absolute or second moments of the distance classes, and m_{2i} are absolute or second moments of the angle classes. NC is the number of lag classes, and f_i^* is a desired number of pairs for a lag class i, while f_i is the actual number of pairs for the network distribution. The placement of points is is iterated so that the f_i values get closer to f_i^* values. Clearly, for the isotropic case, $c = 0$, and the paper concentrates in the case where $b = 0$. An interesting point is that when $a = 0$, $b = 1$, and $c = 0$, the formulation is almost the same as the analysis worked in [Russo, 1984]. The paper goes through five examples. The first and second examples are based on a similar idea of setting $a = \frac{4}{N(N-1)}$, $w_i = 1$, $b = c = 0$. The optimization approach is similar to simulated annealing procedure [Press et al., 1992, Deutsch and Journel, 1997]. The difference between examples one and two is that the first one is a completely Random Sample for the updating step in the optimization, while in example two it is a somewhat regular array. Example three considers $a = 0$ and so the optimization is based on the basis of minimum moments (variances) within the members of the separation distance. Example four is more elaborate, fifty points are chosen at random, with $a = 1$, $b = 0$, and also increasing the number of classes considered to thirty classes of 0-15, 15-30,...,435-450 points. The results for this example are closer to an idealized uniform distribution. The last example is more of a heuristic approach where thirty points are chosen at random and classes are chosen according to direction as well as separation distance. The purpose of this example was to analyze the effect of a more organized sampling procedure. The discussion/conclusions are of great interest for the work reported in this dissertation. Basically, the authors are trying to answer a simple but important question: what is the best scheme for locating sampling sites? This is one of the few papers that tries to specifically approach this issue. Normally, it seems like a large number of points in a random fashion is sufficient for appropriate results. However, the authors say that for geostatistics it is not sufficient to do that. A reasonable criterion may be used to choose sample locations and to assure that the sampling design will enhance the reliability of the sample variogram as an estimator, while constraining

the sample size (number of points). There is also a discussion about the choice of the SS function, whether or not there is a better alternative. The solution does not seem to be simple and probably is dependent on the specific variable measured and the local conditions of the field in consideration. If the only purpose is to meet preset lag class number of pairs, then a pattern can be generated to be very close to meeting the specification, even for large class separation distances. The solution is for points to be concentrated in an area which is equal to the largest lag distance specified. Also, if the maximum class separation is large, then the points will cover essentially most of the field. The most interesting comment about the approach of [Warrick and Myers, 1987] has to do with the last conclusion: a coarse fixed grid (hopefully a regular grid) together with some random points (such as clusters with small transects) show promising and interesting results. The question, of course, cannot have a unique answer and so the field is still open to other suggestions in the future.

The paper by [Rouhani, 1985] addresses the issue of reducing the estimation variance. It is one of the few papers in the literature that poses the basic question of where to sample? and also studies the issue of where is the best place to do further sampling. The paper goes into good detail of analyzing current literature about these two issues. However, the paper really emphasizes much more the issue of the worthiness of adding an extra point to an already existing set of data. In fact, the algorithm proposed by the author relies on the kriging variance, therefore raising the question of what really is the purpose of the paper. The variogram needs to be estimated beforehand, and unfortunately that means this method will hardly be used for a field for which nothing is known.

Inverse Problem

Parameter estimation, also refered to as the Inverse Problem, is the general problem of fitting the parameters of a theoretical model to measured information. The measured information could be from laboratory experiments, or from real world phenomena. The general field of the inverse problem is extremely rich in terms of publications in the literature. Applications

range from physics, economics, engineering and certainly spatial variability. One of the fields that is closely related to variogram models and geostatistical framework is that of the groundwater inverse problem. The groundwater inverse problem has been studied for several decades, as is clear from [Sun, 1994].

The most common approach in the inverse problem is to use a nonlinear regression algorithm to fit a parameter estimation problem. The work described in [Cooley, 1985] is an overview of several methods for solving nonlinear regression groundwater inverse problems. It has a good presentation of the Levenberg-Marquardt method, as well as other methods such as Quasi-Linearization, Quasi-Newton and the Fletcher-Reeves method (which is a form of the conjugate gradient method [Press et al., 1992]). In terms of number of iterations required, the fastest ones were the Marquardt and the quasi-linearization methods, then the quasi-Newton, and finally the Fletcher-Reeves method which did not converge after a maximum allowed number of iterations (100). But when the methods are analyzed in terms of speed per iteration, the slowest were the Marquardt and quasi-linearization methods. The memory requirements were more demanding for the quasi-linearization method, followed somewhat by the Marquardt method, but the one that required the least memory was the quasi-Newton method. The paper then presents the four methods of interest, in particular the Marquardt method is presented as a compromise between the Gauss-Newton method and the conjugate gradient, the way to change its behavior depends on a parameter that essentially weights one method or the other according to the proximity of the optimal solution. The paper then presents the finite difference approach used for the test problems presented in the paper. The solution approach is actually a weighted least squares, where the total weighted sum of errors is minimized with respect to the vector of unknown parameters. The algorithms for the minimization routines are presented in good detail and the notation is consistent and written in a way that the differences among the methods are clear.

Inverse Problem Classification

[Yeh, 1986] presents an excellent Classification of Parameter Identification Methods. The classification used by Yeh is the one presented by Neuman, which divided them into direct and indirect methods. The *Direct Method* treats the model parameters as dependent variables in a formal inverse boundary value problem. The *Indirect Method*, on the other hand, is based upon an output error criterion where an existing estimate of the parameters is iteratively improved until the model response is sufficiently close of the measured output. Another classification method discussed by [Yeh, 1986] uses two distinct categories depending on the error criterion used on the formulation. This classification is consistent with the one given by Neuman, and so [Yeh, 1986] uses this classification to present the methods available in literature:

1. *Equation Error Criterion (Direct Method)* requires that at grid nodes where data are not available that data values are interpolated. The interpolated data contain interpolation errors. Also, if the interpolated data and the observed data (which have noise) are substitued into the governing equation, an error term will result called equation error. The error is then minimized over the choice of parameters. The techniques that are included in this cathegory are (not exhaustively): energy dissipation method, linear programming, multiple objective decision framework, quadratic programming, penalty functions and quadratic programming, and matrix inversion method combined with kriging.

2. *Output Error Criterion (Indirect Method)* is applicable to the situation where the number of observations is limited, and it does not require differentiation of the measured data. One possible disadvantage of this method is that minimization is usually nonlinear and often it is non-convex. A number of special optimization algorithms have been used to solve the minimization problem, and all of them iterate on an initial

set of parameters of the problem, improving them until the observed data gets "sufficiently close" to the computed data. Several techniques borrowed from general control theory have been applied to the problem. The paper offers, by the way, excellent tables summarizing the methods named in each category, what were the parameters estimated, data processing, inverse solution procedure and the reference in literature. The methods worth to mention in the Output Error Criterion are: quasilinearization, minimax and linear programming, maximum principle, optimal control and gradient solution, and of particular interest are Kalman Filter techniques which are common in papers such as [McLaughlin and Townley, 1996] and [McLaughlin, 1995]. The papers and works by Kitanidis ([Hoeksema and Kitanidis, 1985a, Hoeksema and Kitanidis, 1985b, Kitanidis, 1995], for example) are all related to the maximum likelihood estimation and kriging. Several other methods based on gradient searches, linear and quadratic programming, as well as Gauss-Newton methods, Newton-Raphson, and conjugate gradient methods are worth to mention in this category.

Inverse Problem Formulation

[Yeh, 1986] presents a general discussion regarding the formulation of the inverse problem, not only in groundwater but in other relevant fields of hydrology and soil physics. The only drawback is that Yeh emphasizes the groundwater inverse problem so some suggestions or ideas are based on things like the Crank-Nicolson scheme of solution of Finite Elements. In the direct approach (Equation Error Criterion), the method proposed in [Yeh, 1986] is the *Generalized Matrix Method*, taking advantage of a Crank-Nicolson type of approach. The solution of the method is provided by a matrix inversion. For the indirect method, the method suggested is *Gauss-Newton Minimization*. The reason behind using a Gauss-Newton approach is the fact that the Gauss-Newton algorithm has proven to be an effective algorithm to perform minimization. The original and further modifications have been extensively used by several researchers. The method does not require the computation of the Hessian matrix

(second derivatives) which is required by the Newton method. The rate of convergence is better when compared to the classical gradient search procedures. The algorithm works best in unconstrained optimization. Also, if the problem is convex, then the local optimum will be the global optimum and therefore the Gauss-Newton method will work well. However, due to noise in observed data, the problem is usually non-convex and so local minima will be present, which can be a problem for the method. The stepsize of the new iteration is a subject of research. In fact, the method becomes the Levenberg-Marquardt method when the step size becomes shorter as the method approaches (hopefully) a minimum. The Gauss-Newton method algorithm can be modified to accept constraints (such as upper and lower bounds for the parameters) by means of a gradient projection technique.

A good presentation of the jackknife approach can be found in [Lamorey and Jacobson, 1995]. It presents the effects of data sparsity through the analysis of changes in lag increments and in the maximum lag distance. The fitting procedure used employs jackknife kriging and weighted least squares. The authors claim that the weighted least squares method gives good estimates without as much computational demand as, say, maximum likelihood. The two methods are very different, but for a large data set, both methods should give similar estimates of the variogram parameters. There is no mention of the issue of regularization and sample support. The jackknife kriging is used for semivariogram model validation, and the jackknifing error is calculated for specific semivariogram parameters determined from a sample semivariogram fit. The specific jackknife method presented in this paper estimates the semivariogram parameters by minimizing the jackknifing error as a function of the semivariogram parameters. This method estimates the semivariogram parameters without the actual computation of the sample semivariogram, and the procedure goes on by eliminating one data point at a time and then estimating it with the rest of the points, so that the error can be accumulated and used to make a change. This jackknife operator is not really jackknifing according to [Shafer and Varljen, 1990], but it serves the purpose of fitting the variogram. The weighted least squares part is essentially the same procedure as

described in [Cressie, 1993], but using the spherical covariance function, so some algebra is presented for developing a more specific function to be minimized. The method is used with synthetic data generated using the Cholesky decomposition method [Goovaerts, 1997] for creating correlated random fields.

It is worthwhile to mention other techniques of parameter estimation. In particular, *Bayesian Estimation Methods*, which incorporate prior information into the parameter estimation process. Essentially, mean and correlations of the parameters are known and have been measured in the field. Among these methods is the Composite Objective Function, as well as the *Kalman filter*. The application of this last technique, requires expressing the model of interest in terms of a state-space formulation that consists of a vector state equation and a vector observation equation. Parameter estimation requires the augmentation of the state vector to include the parameter vector as another state variable. Assuming the errors in the state and observation equations have zero mean and are gaussian with known covariance matrices, then the Kalman filtering method can be applied for simultaneous, recursive state, and parameter estimation. The method clearly requires prior information and so it is classified under bayesian estimation techniques. As a final comment, the primary purpose of incorporating prior information into the inverse problem is to reduce uncertainty, not to improve the model fit. In fact, in most cases the prior information usually worsens the fit.

Incorporating prior information as a penalty function does not affect the feasible region of minimization. Careful use of prior information will produce stable and reliable parameter estimates. A good example of Bayesian estimation applied to a more complicated problem that estimates moments conditioned to data is [Graham and McLaughlin, 1991]. The moments of the property of interest are conditioned to observed data, under the assumption of non-stationary fields, using Bayesian estimation theory, combined with data conditioning, so that all allowable information is used. The quantities that are estimated are the mean and the covariance. [Graham and McLaughlin, 1991] use and present the theory for a Kalman filter, with a non-fickian model, and with a solution using finite elements with a

double grid. Data from a tracer test at the Borden aquifer in Canada are used. The main purpose of the paper is to evaluate the performance of the stochastic model developed. The basic reasoning behind the model is that spatial variability in the hydraulic conductivity induces variability in a steady state velocity field, which then induces variability in solute concentrations. There are analytical expressions for the unconditional ensemble moments of velocity, which are derived from the unconditional ensemble moments of log hydraulic conductivity. The result is that the velocity moments are related to the ensemble moments of concentration by a set of coupled partial differential equations. A double grid finite element solution is used to obtain the values of the concentration mean and covariance, as well as the cross covariance between concentration and velocity. The macrodispersion was assumed non-Fickian. With the results of the nonstationary concentration covariances obtained with the FEM solution, the estimates of velocity and concentration are conditioned on a small set of concentration measurements. The analysis of the residuals of the model is consistent with the model's own estimates of concentration uncertainty. One of the main reasons to have used Bayesian estimation is because it provides a particularly convenient way to integrate model predictions and field data. The model is treated as a source for *a priori* information, and then when new field observations become available the model can be updated or conditioned to this new information. The ultimate source in concentration variability is the assumed spatial variability in hydraulic conductivity. Since the model does not assume that field-scale transport is Fickian, the macrodispersive flux is treated as an unknown. The flux is obtained from a closure equation which depends on the velocity mean and covariance, and the mean concentration gradient. The conditional covariances computed by the model provide a convenient basis for designing field sampling programs and monitoring networks. Conditioning integrates modeling with data collection, and provides a systematic conceptual framework for carrying out a site investigation.

The concept of measurement conditioning, according to [Graham and McLaughlin, 1991], combining physically based models with field data is an excellent and useful tool for many

water resource applications. The conditional mean and covariance can be approximated by a nonlinear estimation algorithm known as the *extended Kalman filter*. This algorithm can be put in two parts:

1. A set of partial differential equations which propagate velocity and concentration moments forward from measurement time t_n to measurement time $t_{n+1} (n = 0, 1, 2, \cdots)$. This part performs a function similar to a traditional groundwater model.

2. A set of algebraic equations which condition the propagated moments on measurements collected at t_{n+1}. This part performs a function similar to a traditional estimation algorithm.

The two steps mentioned above, moment propagation and updating are repeated whenever new measurements become available. The Kalman filtering algorithm is a convenient way to unify the two processes so important in groundwater modeling. The specific equations that describe the Kalman filtering algorithm are presented in [Graham and McLaughlin, 1991] and will not be presented here. The algorithm yields exact conditional moments only when the state equations are linear in the estimated variables and when all sources of uncertainty are normally distributed. In the case of groundwater modeling this is not the case because velocity and concentration (which are simultaneously estimated) appear as a product in the state equations. The algorithm yields better results as the hydraulic conductivity variability is moderate and the measurements are abundant. There is a finer grid (with N nodes) and a coarse grid (with M nodes). The coarse grid nodes are a subset of the fine grid. If M is significantly smaller than N, the dual-grid approach provides a large reduction of computational effort, as compared to traditional solutions of the undecomposed covariance equations. Since the finite element equation is still dependent on the fine grid, the fine grid resolution's advantages are retained. A Monte Carlo analysis of the matrix solution using Cholesky decomposition was also performed. It was found that $\Delta_l = \Delta_t = \lambda_f$ for the coarse grid spacing, and $\Delta_l = 20\ \alpha_l$ and $\Delta_t = 20\ \alpha_t$ for the fine grid. λ_f is the log

hydraulic conductivity correlation scale, α_l and α_t are the pore-scale dispersivities, the l and t subscripts refer to the directions along and normal to the mean flow, respectively. First of all, moment conditioning is more sensitive to the quality of the prior estimated mean plume, and the resulting concentration perturbations at measurement points. A key to the conditioning process is to define the site-specific spatial variability of the groundwater velocity field. The model developed is a great tool because it:

1. Provides a predicition of the extent of the solute plume.

2. Provides an estimate of the reliability of this prediction.

3. Uses the Bayesian estimation concepts of measurement conditioning, which is used to combine a physically based model and field data.

It is then clear that in the parameter estimation problem in any discipline, but in particular groundwater and hydrology, incorporating prior information is always desirable. The effect of this is to reduce uncertainty in the estimates of the model that is being fit, but it might not improve the fitting of the model itself. A more likely outcome is that incorporating more prior information will worsen the fit of the model itself. In any case, the parameter estimation problem in general has been studied well enough but there is no indication anywhere in the literature that the issue of sample support for variogram estimation has been addressed before.

Inverse Problem using Field Data

Parameter estimation is also an important issue when using field measured data to assess important characteristics and draw conclusions from field experiments. An important example of this type of use is found in [Ellsworth et al., 1996]. This paper talks about sampling schemes, parameter estimation and model discrimination, presenting two sampling methods in field experiments: soil coring and solution samplers. The paper has three main objectives:

1. Compare the efficacy of two sampling methods: coring and solution samplers.

2. Study the variation of transport model parameters with increasing depth of solute leaching.

3. Perform model discrimination to examine the transport process operative within a field plot.

Bromide, nitrate, and chloride were applied sequentially to a plot instrumented with several solution samplers (12 of them, at each of two depths). At the end of the experiment, the field was destructively sampled, 2 meters by 2 meters by 2 meters. The mass recovery for solution samplers was of the order of 63% to 83%, whereas the mass recovery for the coring procedure was about 95% to 105%. The mean solute velocity was clearly less for solution sampler data than for coring. However, solution samplers and coring provided similar results for vertical dispersion, showing that the dispersivity increased linearly with mean residence time. The depth profiles were well described with a stochastic convective lognormal transfer function model (CLT). This paper is in a way picking up where the paper by [Jury and Sposito, 1985] left off. The idea is once again to compare solution samplers and coring as a way to do mass recovery. Also, the paper justifies the need for more solute transport field experiments to compare to the theoretical models and programs developed for several years already. The issue of scale-dependent dispersion is studied, and a typical idea is to fit the mean solute velocity, V, and the dispersion coefficient, D, in the CDE model to tracer data sampled sequentially in time, and then plot the dispersivity, α (ratio of D/V), against time of sampling (or mean travel distance). Generally, α is observed to increase with any of these. However, there is not a general consensus of scale-dependent dispersion in solute transport experiments through unsaturated field soils. Some literature shows a linear dependence, others show a non-linear relation, while others show an erratic behavior. The specific results of this paper are of a linear dependence. The sequential application of solute tracers in a series of pulses, with different sampling methods provides a reasonably good tool

for process and parameter identification.

Examples of Geostatistical Applications

The theory of geostatistics, scale, and parameter estimation have been applied in numerous cases. The literature is full of examples and a sample of that material will be presented here. Satellite images are a good example where geostatistics can be successfully applied. In fact, the book [Cressie, 1993] includes a large amount of applications in the field of remote sensing. Another example is [Brutsaert, 1998]. This paper provides important observations about geostatistics and variograms, such as a definition of nugget effect as the variability that cannot be resolved by the available data sampling spacing. The variable of interest was the normalized difference vegetation index (NDVI), measured by the SPOT satellite and the Landsat satellite, from a grid of $667x667$ pixels measurements with $30m$ resolution (representing and area of $20x20Km$). The paper presents explanations on the effect of making measurements beyond the correlation range distance (which turned out to be in the order of $900m$), and also the effects on the variogram from small scale measurements. The last remarks about geostatistics by the author deal with the issue of multiple spatial variability structures at different scales, and how to represent them with nested structures. It is not a complicated paper, it shows how the concept of spatial variability can be applied to some areas of remote sensing.

Two good papers by Dagan show a more complicated aspect of geostatistical applications with a more elaborate mathematical formulation. The first one, [Dagan, 1982], studies the direct groundwater problem, i.e., just plain groundwater modeling. But the author considers the formation properties (hydraulic conductivity, transmissivity) as well as flow variables (head, specific discharge, solute concentration), and all are treated as random variables and so are assumed known with uncertainty. The main purpose of the paper is to study the influence of conditional probability of input variables upon the statistical structure of the dependent variables. The unconditional probability density functions are supposed to be

stationary multivariate normal, while conditioning accounts for the measured values at a few points of the formations. In unconditional modeling, the author models the variable of interest (conductivity, whatever) as stationary and represented by its constant mean and variogram. The ensemble of formations on which statistical calculations are carried out represents all aquifers with the same probability density distributions. By conditional modeling, the author means that the measured values at a few points are kept fixed and uncertainty is only present at any other points. Thus, statistical computations are performed for the subensemble of aquifers which preserve the measured values, and as a result, both input and output variables are nonstationary. The second paper of this series is [Dagan, 1985]. This paper appeared 3 years after publication of the original direct problem solution paper. The paper is more simplistic than part 1 of the series in the sense that the author defines the inverse problem as this: to determine the value of the transmissivity at various points, given the shape and boundary of the aquifer and recharge intensity, and given a set of measured log transmissivity and head values at a few points. The log transmissivity distribution is regarded as a realization of a random function of normal and stationary unconditional probability density function. The aim of the inverse problem is to estimate the conditional normal probability density function of the log transmissivity values, conditioned on the measurements of head and log transmissivity, which is expressed as the unconditional joint probability density function of the log transmissivity and the head. The problem is, bottom line, to determine the unconditional head - log transmissivity covariance and the head variogram for a selected log transmissivity covariance which depends on a few unknown parameters. The way to solve this is by a first-order approximation of the flow equations (very mathematically intensive, like any other paper by Dagan). The paper uses the exponential model for an example worked out by the author.

There are applications of the whole idea of conditioning to the solution of the flow problem, both a direct and inverse problem, such as [Dagan, 1982] and [Dagan, 1985]. In the first paper, Dagan presents a direct problem approach to 1, 2, and 3-Dimensional flow problems

toward a well, in a stochastic framework. The solution applied was the perturbation method, where one of the restrictions is that the variance of the variable (for example, log hydraulic conductivity) be small. For the case of 1-Dimensional flow with average uniform head gradient and unconditional modeling, the head variance is not constant, but the variogram is stationary. When recharge is present, even the variogram is not stationary. Conditioning the hydraulic conductivity at a point results in spatial variation of the head gradient and reduction of head variance everywhere. For the 2-Dimensional flow, and with uniform average flow the head field is also not stationary although the variogram is stationary. But the most interesting case is nonuniform flow toward a well of given discharge, the head variance at the well is quite large and conditioning of transmissivity at the well reduces this variance considerably. The 3-Dimensional case considered was quite simplified (because it is inherently very complicated). The head field in this case is stationary for unconditional log conductivity pdf, and conditioning in this case leads to a local reduction of the head variance. An interesting conclusion by the author is that, in general, solving the direct problem of steady flow through heterogeneous aquifer would consist of the following steps:

1. Using measurements of transmissivity (either field measurements, or inverse problem results), estimate the trend and a stationary variogram of the logarithm values. The simplest case is the one of a constant mean.

2. Using a conditional multivariate normal distribution, the conditional mean and covariance matrix of the transmissivity at an arbitrary set of points can be derived.

3. Given the shape of the flow domain, appropriate boundary conditions, and recharge, solve the direct problem by Monte Carlo conditional simulations and numerical methods. To do this, generate realizations of the transmissivity field using the conditional multivariate distributions. The idea is that each one of those realizations is deterministic, therefore one can apply any existing model/code to solve the groundwater

problem. The only thing to keep in mind is that the heterogeneous field is the conditional realization. So, by solving several realizations of the conditioned field as the input field in the groundwater code, the entire statistical structure of the head and specific discharge can be determined.

4. A good detour in numerical complexity can be taken when solving the direct problem using a first-order perturbation approximation, as long as small values of the log transmissivity variance are present. The scheme is developed and suggested in good detail in the paper, so I am not writing it here. It uses the Green function for the given flow domain.

The second paper of the series, [Dagan, 1985], as in the first part, used a systematic approach to study the effect of head and transmissivity measurements upon the reduction of the uncertainty of log transmissivity. The main conclusion is that the impact of the measurements of the head is not as important as the measurements of log transmissivity, but then a good design and combination of the two types of measurements should do a significant reduction of the variance of log transmissivity. The conclusions drawn by the authors are quite interesting. In the conclusions, it's interesting to note that the authors outline the assumptions and steps to follow for the use of the method proposed:

1. The log transmissivity spatial distribution in the actual aquifer is supposed to be a realization of a normal stationary random function.

2. The probability distribution function of the log transmissivity solution of the inverse problem is the conditional pdf conditioned on the measured values of log transmissivity and head. The latter pdf is expressed with the aid of the unconditional pdfs by Bayes formula.

3. The unconditional pdf are derived by solving the direct problem along the line of

[Kitanidis, 1995]: the covariance of the log transmissivty is supposed to have an analytical expression and to depend on a few unknown parameters, and the mixed covariance of the head and the log transmissivity, as well as the covariance of the head alone are derived as functions of the same parameters, by solving the direct problem. The parameters should be obtained by a Maximum Likelihood procedure, using the measured values of head and log transmissivity.

4. To make things simpler, a few extra assumptions are adopted. The expectation values of log transmissivity and of the gradient of the head are assumed to be constant. For the solution of the flow equations, a first order approximation is made (valid as long as the variance of the log transmissivity is small), and the flow domain is assumed to be infinite.

The authors gave an example using an isotropic, exponential model for the covariance of the log transmissivity. One of the most important results is that the method can be used to estimate the necessary distance between measurement points in order to achieve a given reduction of variance in the log transmissivity. Thus this is a quick solution to the inverse problem in this way. There is a further way to make it quicker and it is to choose a neighborhood around the point of interest, containing a sufficiently large number of measurement points, considered apart from the rest of the measurements. Making the assumptions given in step 4 above gives then a quick solution to the inverse problem. A more elaborate scheme is using a mixed numerical-analytical procedure, following this idea:

- The expectation of the head is determined by solving numerically the flow equation as described in the paper, using the specified boundary conditions, and determining the expectation of the log transmissivity from field observations.

- The inverse problem at a point is solved in a local neighborhood by the quick method outlined before, using the expectation of the log transmissivity and $\mathcal{J}$ (see the paper) from the numerical solution.

The procedure can be used for the solution of the inverse problem, or more interestingly, to estimate the worth of extra measurements regarding reduction of the uncertainty or improving models depending on the log transmissivity. It should be clear at this point the importance of accurate field measurements for a successful data conditioning. The importance of field experiments is noted in several references, in particular [Jury and Sposito, 1985] and [Ellsworth et al., 1996], specially if coring is the choice for sampling the soil.

Applications of Kriging and Co-Kriging

The next series of papers are good applications of geostatistics but on the line of kriging and co-kriging. The first one is in [Stein et al., 1988], which talks about using kriging and co-kriging as ways of improving MD-30 mapsin the Netherlands. Observation patterns, i.e., sampling design, is an idea treated in this paper. Sampling stratification is presented as a way to reduce the uncertainty in spatial variability, as well as using existing knowledge of soil maps and geological maps and descriptions to further improve the results of point data on moisture deficits. Also, the paper suggests the use of previous terrain surveys and common sense/knowledge along with remote sensing and any other available information to improve the mapping results. The precision of the MD-30 maps was measured in terms of two quantities: the mean squared error, and the mean variance of the prediction error. For unstratified maps, the standard deviation of the prediction error largely depends on the observation pattern (sampling). But with stratification, the precision of the predicted values increased in strata with low MD-30 variability and also an apparent decrease in strata with high MD-30 variability. The use of co-kriging increased on average the precision of MD-30 maps by about 10%. The paper is a clear example of the benefits and improvements of the use of soil-survey information and in general of any prior information about the site to improve the precision of predictions when using kriging and co-kriging.

The second paper is [Petach et al., 1991], which shows a study using the computer program LEACHM, combined with a GIS model for pollution in soil and groundwater, allowing

for some statistical variability to be incorporated. The paper makes use of pedotransfer functions, which correlate soil physical properties and characteristics to the mean and variance of the hydraulic properties. The paper also incorporates in the analysis a Monte-Carlo type of approach to study the variability of parameters used in the model. This paper mentions, as other papers do, the fact that many environmental regulatory policies are based on computer simulations. Care should be taken on the appropriate use of such computer models, in particular the laws that regulate the rate of application of pesticides.

The last paper in this series is more of a comparison of a geostatistical approach versus a non-statistical approach. [Tiktak et al., 1998] presents a study of cadmium transport modeling, using a process-oriented model SOACAS (fate of a chemical, completely mixed compartment, Freundlich isoterm sorption). Also used a is GAM approach (General Additive Modeling), which is a regression model of cadmium modeling with locally- weighted smoother. This GAM approach is available in statistical packages such as S-Plus. For model discrimination, the authors used the AIC (Akaike Information Criterion). The paper is from the Netherlands, regarding Dutch policies and other issues about land use and environmental pollution in that area. The results of the procedure used in this paper are promising; using very few assumptions, the authors were able to obtain a realistic geographic pattern of cadmium concentration. The work could be extended to predict future metal contamination distribution given emission/input values. Cadmium is of a main concern because of an incident that happened in Japan years ago where paddy fields were polluted with cadmium from an upstream mine. In a country such as the Netherlands, the sustainable soil quality management agency imposed a ban on any further cadmium pollution. To study trace metal behavior, there are a number of programs/models available, such as CHARON, ECOSAT, and PHREEQM. The problem with these models is the data input requirements for a large scale application and the computational burden are both large. [Tiktak et al., 1998] used SOACAS because it is a simple dynamic model, and it is aimed at the prediction of accumulation of trace metals in the soil. The authors developed SOACAS for this project, and

present a technique for model validation at a regional scale. The procedure is called hind-cast simulation, which consists of testing whether it is possible to reconstruct the soil's present state using historical data on metal loadings. The problem with regional-scale models is that there is not really a reliable way to test the physical processes and assumptions in the model itself. For that reason, the authors did some field experiments at the plot scale to test the validity of the physical principles in it. However, as is pointed out, this is not the ideal situation, but is hard to improve on because the cost of a regional scale experiment would be prohibitive. The authors used about 2500+ measurements of cadmium contents taken between 1960 and 1988. As in the applicability of any other model, it is extremely important to make clear the scale of available and simulated state variables. The model developed by the authors predicts gridcells of $500x500m^2$, but the observations are of variable size. Thus, the authors used block-kriging as a scale-changing interpolator. The authors did not stratify the survey area (they argue that if they had done so, the computation of the experimental variograms would decrease in quality because the number of pairs is reduced). Rather, they use a locally-weighted smoother in combination with a regression model within the framework of Generalized Additive Modelling (GAM) as an alternative to block kriging. GAM's are extensions of Generalized Linear Models (GLM). The idea in the GAM framework is that the models are flexible in that they allow for parametric and non-parametric model responses in one single model. They include smooth functions (smoothers), and as such a smoother displays a trend in the data without forcing them into a rigid form of dependence (which is different from the approach of say, Kitanidis). These functions can fit any possible shape of response curve dependent on the responses observed at neighboring values. Even smooth surfaces can be included into the model. The response curve is data-driven rather than model-driven (which can be a disadvantage). Mathematically, the GAM model is represented by $g(\mu) = \alpha + \sum_{j=1}^{p} f_j(X_j)$, where $g(\mu)$ is called the link function, α is the intercept, and $f_j(X_j)$, for all $i = 1, ..., p$, are functions of the p predictor variables. These functions can be linear, polynomial, or smoothers. The link-function describes in which way

the estimated values μ_j are related to the right-hand side of the equation. The link function can be the identity function ($f(x) = x$) or it could be as complicated as the logarithm of another function. The logarithm is a common transformation to turn a multiplicative model into an additive model. The specific details of the applicability of the GAM framework to this study are in the original paper and will not be presented here, except to say it is a very interesting approach to solve a problem to block-kriging.

The final part of this literature review presents two papers that show future directions and general new ideas that generalize many of the ideas in the recent literature. The first one is in [McLaughlin, 1995] which talks about data assimilation as a separate science. Data assimilation is a generalization of objective analysis which provides time-dependent spatially distributed estimates that can be updated whenever new data become available. This paper reviews the current state of the art in data assimilation (which originated in the areas of meterology and geophysics and petroleum exploration). It goes over the methodology, current ideas, the forward and inverse problems as particular cases of data assimilation and explains in detail why most of the simulation and estimation schemes (including cokriging) are special cases of the Kalman filter. It is a very good, down to earth paper that goes into good detail in issues such as geostatistics, and basically tries to give some history into the data assimilation problem in hydrology, which is also discussed in [Chiles and Delfiner, 1999]. The need for new and more extensive data sources, says [McLaughlin, 1995], will make a good improvement in the area of hydrology, getting it up to speed to geophysics, seismology, petroleum engineering and so on. A big point in the conclusions is the fact that for a more efficient data assimilation it is necessary to have a better description of temporal and spatial variability.

The second one is [Beckie, 1996]. This paper presents very interesting conclusions that deserve to be mentioned: the main issue is that the scale of a single parameter does not simply depend upon a support volume, but rather upon the model, model conditions, and inverse methods used to measure the parameter. As a matter of fact, model parameters

represent the effect of physics that are not explicitly resolved in the model upon physics that are explicitly resolved. Model parameters are thus dependent upon the scale of the model that defines them and the nature of the resolved and unresolved physics that the model describes. The paper basically presents a spatial filter to represent the relationship between parameters measured with different techniques. Another spatial filter is used to represent the smoothing effect that occurs when a network of discrete samples is interpolated using geostatistical approaches such as kriging. Most of the conclusions drawn in this paper are: a network of measurements can only resolve the large scale component of the parameter field. The subgrid component of the parameter field is not resolved. The subgrid component can only vary in the bounds of the network scale and the sample support. The unobserved subgrid component of the parameter field is responsible for the "closure" problem and other errors. The closure problem arises when the subgrid scales are not properly represented by the model, and it is a difficult problem to address because of the fact that subgrid scales are not observed with measurements taken on a network. Near the end of the paper it is stated that low scale resolution can be damped by increasing the support size, and also that low scale resolution can be a problem that has to be dealt with carefully.

Monte Carlo Anlysis

[Wagner and Gorelick, 1986] used a Monte Carlo Analysis for the computation of the statistical properties of the parameters and functions used in their particular analysis of variability and sensibility. In the paper, the authors used a groundwater problem for their research purpose. The statistical analysis was performed using a Monte Carlo approach. The specific characteristics of their approach are that they picked up previous work for a 1-D case and essentially wanted to evaluate the reliability of the parameter estimation problem. The groundwater problem of interest is that of solute transport, and three hypothetical examples are presented. The first one represents a 1-D column where the data have been tainted with

artificial random errors and the idea is to show the sensitivity of the parameters to such errors. The second example deals with the estimation of dispersivity and effective porosity of an unsaturated soil using concentration data from column tracer experiments, and the data are compared with synthetically created data. The third example is related to parameter estimation of advective and dispersive transport with transient storage, which is described by a second linked partial differential equation.

The work by [Parker and Albrecht, 1987] analyzed saturated hydraulic conductivity, solute dispersivity and sample support effects through a Monte Carlo approach. The authors used a multilognormal joint pdf of conductivities and dispersivities and an ensemble model for a set of "noninteracting vertical solute stream tubes". Two Monte Carlo approaches were used: a two-step Monte Carlo analysis, and a one-step Monte Carlo, followed by a first order uncertainty analysis. The research was done by measuring saturated hydraulic conductivities and solute dispersivities under saturated flow conditions in the laboratory for three different sample volumes: 92, 471 and 1770ml. The samples were taken at two depths (in two different soil layers) along closely spaced parallel transects. The variances of the log of the hydraulic conductivity decreased and the mean dispersivity increased with increasing core volume, probably due to greater within-core and lower between-core velocity variations for larger cores. The mean of the hydraulic conductivity was significantly lower for the smallest core size suggesting structural disturbance. The Monte Carlo approach was used in two ways: the first was to investigate the effects of sample volume on field scale transport predictions, and the second was to evaluate effects of uncertainty in parameter estimates. For the second usage, a second level of Monte Carlo analysis was performed with a normally distributed parameter assumption. The differences between beakthrough curves predicted for each core size were relatively small and within the 95% confidence interval limits.

Expressing Uncertainty with Confidence Intervals

The use of confidence intervals for expressing uncertainty is a common practice in the scientific literature. Several papers and authors address the importance of such approach. [Knopman and Voss, 1987] deals with the sensitivities of parameters of a given fitting routine, to the data error. The authors do an analysis based on regression, defining sensitivity basically as a partial derivative of the variable/parameter of interest with respect to some quantity of interest. The analysis encompasses both spatial and temporal sensitivities in solute transport in porous media, using the advective-dispersive equation. Although this paper is related to transport in porous media, there are some interesting points. For instance, the sensitivity of the dispersion coefficient to spatial and temporal sample location is usually at least one order of magnitude less than the sensitivity to the velocity. Also, by assuming the probability distribution of random error in observations of solute concentration one determines the sensitivities. Another important observation is that if a sampling design minimizes the sensitivity of some output variable to one parameter, it does not mean that it will reduce the output variable's sensitivity to other variables.

The section on parameter estimation clearly presents the fact that no matter how calibration or parameter estimation is done, it will always be an optimization problem at heart. Non-linear regression has been the standard procedure to do parameter estimation, not only in groundwater hydraulics but in other fields as well. This paper presents the estimation process in a simplified form:

$$F_{obs} = F_{com} + E_r + E_s \tag{1.2}$$

Where the term E_r is the random error in observations, and the term E_s is the systematic error due to an incorrect physical model. The terms F_{obs} and F_{com} are respectively the observations and the model predictions of the property of interest. In simple words, the equation says that if we have a model that represents a physical phenomenon for which we have observed values, the observations will equal the model predictions plus an error

composed of two parts: incorrect model predictions plus a random error in the observations. Usually the model contributes to the error because it is an over-simplified model which does not properly account for all the processes or components that affect the observations. In practical terms, it is very difficult to completely separate the two error terms, and so a simplification must be made. Normally this simplification is to assume that $E_s = 0$, i.e., to assume that the physical model is correct. Also, the random error is assumed to be normally distributed with zero mean. The random error of observations, E_r, is changed to an E term, and so what we end up having is: $E = F_{obs} - F_{com}$. If E is represented as a $N_{par} x N_{par}$ matrix, then the parameter estimation problem becomes an exercise in minimizing: $E^T W E$ where W is a weighting matrix related to the estimated variance in the observations. The method to solve the system can be any of the methods in the literature, for example the methods mentioned in [Yeh, 1986] and [Press et al., 1992]. At the end of the iterations, the theoretical covariance matrix of the parameters is basically obtained by using the last iteration of the optimization process (if using a Newton-like method), and takes the following form:

$$Cov = (E^T W E)^{-1} \cdot \sum_{i=1}^{N_{obs}} (F_{obs_i} - F_{com_i})^2 / (N_{obs} - N_{par}) \tag{1.3}$$

Where N_{obs} is the number of observed values, N_{par} is the number of parameters being fitted, F_{obs_i} represents the observed values, F_{com_i} is the corresponding value predicted by the model, W is a weighting matrix as mentioned before, and E is the error matrix explained before.

Sampling design is another optimization problem. [Knopman and Voss, 1987] say that it is an iterative process where one cannot simply state a design and then claim that it is optimal. A criterion is presented which is in a way similar to the covariance matrix. The authors present the D optimality criterion, based on the covariance matrix, where one minimizes: $(E^T E)^{-1}$. This matrix can be derived theoretically or computed by a method such as the Monte Carlo approaches.

Some approaches that use confidence intervals are in [Bardsley et al., 1985], where an inverse problem approach is applied to groundwater. An excellent study of confidence intervals under change of support situations is presented in [Tercan and Dowd, 1993], where

the authors use Indicator Kriging, Probability Kriging, and Class Indicator Kriging under a change of support. The treatment of the change in the histograms is done using both the Affine correction and the Lognormal Transformation (and, by the way, gives essentially the same results if the data are normally distributed). The notation is a bit confusing, but after appropriate transformations to the notation in [Isaaks and Srivastava, 1989] it is more clear what the authors did. The analysis is performed with synthetic data generated using the turning bands method [Deutsch and Journel, 1997] with a spherical covariance function. [Tercan and Dowd, 1993] use three data sets, the first one is the simulated data itself and the other two are transformations of the first one, specifically to make the data be log normal distributed in one case and gamma distributed in the other case. The authors did not analyze any regularized variograms as they were only concerned with the change of support for the resulting histograms. In the conclusions, the authors say that a change of support always increases the differences between observed and computed values. Cumulative indicator kriging yielded the best results to minimize this difference, and gave smaller confidence intervals, while class indicator kriging gave wider confidence intervals, which can be explained as a natural consequence of a high relative nugget effect associated with the class indicators.

The work in [Shafer and Varljen, 1990] used the jackknife variance approach:

$$\sigma_J^2(h) = \frac{1}{g(g-1)} \sum_{j=1}^{g} \{J_j[\gamma(h)] - J[\gamma(h)]\}^2 \tag{1.4}$$

Where g is the number of partitions of the entire dataset. The values of $J_j[\gamma(h)]$ are the variances of each of the partitions, and $J[\gamma(h)]$ is the mean value of the jackknife variance. From this expression we can use $\sigma_J(h)$, the root mean square error of estimation, to estimate the confidence limits on $J[\gamma(h)]$. If we assume that the errors are normally distributed with zero mean, then the upper and lower bounds of the estimates are represented by:

$$\bar{J} \pm 2 \cdot \sigma_J(h). \tag{1.5}$$

This paper then shows more applications of the Monte Carlo type of approach using data synthetically generated by the turning bands method [Deutsch and Journel, 1997]. The number

of partitions, g was analyzed using a Monte Carlo approach, where the purpose was to find the minimum size of the field to give a stable variance. The results of the paper claim that the estimates of the jackknife variance are within the 95% limits obtained by the Monte Carlo analyses. After the validity of the jackknife method was demonstrated, then the authors proceed to use real field data of pesticides and fertilizers in groundwater from 1982. The drift was neglected because it was fairly small compared to the drift of the nitrate-nitrogen concentrations that were used for the jackknife and Monte Carlo analyses. To the present time, there is, in the literature, no comprehensive study of regularized variograms and their application to estimate point variograms by means of Monte Carlo analyses and thus using confidence intervals. Confidence intervals are a viable alternative that has proven to be valid and sound elsewhere in the literature. The next sections will further discuss the use of confidence intervals and show by means of real applications their potential of these techniques.

1.3 Objectives

The general purpose of this dissertation is to study the effect on the sample support size on the estimation of the point variogram. What exactly should be addressed then? This can be summarized in the following items:

- Study the effect of sample support on the estimation of the spatial variability structure. What are the effects of sample support size on estimates of variability, robustness and uncertainty?

- Investigate the impact of sample support on the variability of the estimates of the point variogram at different lag separations. In other words, what should be expected regarding the confidence intervals of the estimates of the variogram function values and regarding the variogram model parameters themselves?

- Analyze the histogram of the data and the possible transformations that happen when regularization and sampling are performed.

- Study, develop and apply criteria to discriminate and select synthetic data generated with conditional simulations from field measured data.

- Offer ways to explain and recommend sampling strategies. The basic question that want to answer is: given a real field for which very little information is known; how should it be sampled?

These issues are frequent questions in the literature (see for example [Addiscott, 1993]). The next group of objectives address how to answer these questions. The questions do not have unique answers; rather, the approach followed in this dissertation entails the following steps:

- Develop computational tools to estimate the point variogram given information in several formats: regularized variograms, mixed support variograms, point variograms and dispersion variance measurements.

- Use the computational tools developed and test their efficacy and applicability in real problems. Two different data sets are available, one is based in the results and field data already presented in [Ellsworth and Boast, 1996] and the second set of data, from Williams Field near Champaign, IL, has been partially presented in [Warrick et al., 1999]. For this data set, conditional simulations on Phosphorous and CEC data were performed.

- Use the program BIGAUS and indicator variograms, together with probability plots to test if the gaussian assumption is correct in order to determine an appropriate method to generate conditional simulations.

- Use a Monte Carlo type of procedure to estimate the 90% and 95% level confidence intervals of $\gamma(h)$ as a function of h. From the same analysis, estimate the same confidence

level intervals for histograms and apply the appropriate sample support corrections. Estimate the variability of the variogram model parameters by traditional statistical analyses and confidence interval analyses.

- Use the same Monte Carlo analysis from three numerical experiments to estimate the sensitivity of the variogram parameters and the variogram function itself to changes in sample support size and also to variations in the sampling patterns.

With the previous objectives, what are the primary conclusions to derive? These can be summarized as follows:

- It is common sense that using more sampling points improves the point variogram estimate and thus makes the size of the confidence interval narrower. But what is the quantitative relationship between number of sampling points and the confidence intervals? What types of information are most useful in the variogram fitting procedure?

- Also, we want to make an assessment of how useful it is to combine various types of information in the parameter estimation routine. This will be quantified using the analysis of confidence intervals.

- We also want to determine the effect of sample support on the histogram. There are several techniques but the literature is very clear in saying that, except disjunctive kriging [Chiles and Delfiner, 1999], none of the methods to assess the change in histogram has a solid statistical basis. In addition, most of the techniques assume that the variance reduction factor is known (i.e., in the process of going from point information to regularized information).

- Conclude what kind of information has a more marked effect on the parameter estimation procedure. Basically, from the Monte Carlo analysis we want to interpret what sampling pattern design as well as what sample support size should be a good recommendation.

Chapter 2

Theory and Methodology

2.1 Assumptions

The first and probably most important assumption in this dissertation is that the variables we are analyzing are second order stationary [Journel and Huijbregts, 1978]. The meaning of first order stationarity is that the first moment, i.e., the mean, exists and does not depend on the support point (location). Thus the mean is constant throughout the field. The assumption of second order stationarity implies, in addition, that for each pair of regionalized variables, the covariance exists and depends only on the separation vector, h, of the two variables. In this work, we also restrict ourselves to isotropy, therefore the vector h is equivalent to a separation distance. This statement is very important in the definition of terms and variables. Another important assumption that is implied by the second order stationarity is that of intrinsic hypothesis. This assumption says that the mean exists and does not depend on the support point x, and also given random function $Z(x)$, for any given distance h the increment $Z(x+h) - Z(x)$ has a finite variance which does not depend on x. Thus the intrinsic hypothesis can be seen as the limitation of second order stationarity to the increments of the random function $Z(x)$. And the last assumption considered in this work is that of ergodicity: a random function $Z(x)$ is said to be ergodic in a parameter p if the corresponding realization statistics of $\hat{p}$ tends toward p as the size of the field increases.

2.2 Concepts and Definitions

A regularized variable is a variable that is spatially-averaged over a certain support volume V, centered at a location x in space. The support volume has a meaning of either a line segment, an area, or a volume, depending on whether the dimensions of the problem are one, two, or three. The regularized variable is the mean value of the point variable:

$$Z_V(x) = \frac{1}{V} \int_{V(x)} Z(s)\, ds \tag{2.1}$$

For a random variable $Z(x)$, where x is a location in space, the point variogram is defined as the expectation

$$\gamma(h) = \frac{1}{2} E\{[Z(x) - Z(x+h)]^2\} \tag{2.2}$$

where h is the separation distance between two points. Note that this definition assumes the intrinsic hypothesis to be true. If instead of using the random variable $Z(x)$ we use a regularization of the variable, we get then a regularized variogram:

$$\gamma_V(h) = \frac{1}{2} E\{[Z_V(x) - Z_V(x+h)]^2\} = \frac{1}{2} \sigma_e^2[V(x), V(x+h)] \tag{2.3}$$

A more general equation, the mixed-support variogram, can be computed when instead of one regularized variable, we have the expectation of the same variable at two different volumes of regularization. In such case, we have:

$$\gamma_{Vv}(h) = \frac{1}{2} E\{[Z_v(x) - Z_V(x+h)]^2\} = \frac{1}{2} \sigma_e^2[v(x), V(x+h)] \tag{2.4}$$

In the last two definitions, the function σ_e^2 is referred to as the estimation variance. In general, the estimation variance is defined as:

$$\sigma_e^2[v(x), V(x+h)] = 2\,\bar{\gamma}[v(x), V(x+h)] - \bar{\gamma}[v(x), v(x)] - \bar{\gamma}[V(x), V(x)] \tag{2.5}$$

The last equation is defined in terms of a function $\bar{\gamma}[v(x), V(x+h)]$, which is the mean value of the function $\gamma(h)$ (i.e., point variogram) when at one of the two points separated by a

distance h the support volume is V and at the other point the support volume is v. In this dissertation, we use the following definition:

$$\bar{\gamma}[v(x), V(x+h)] = \frac{1}{vV} \int_{v(x)} \int_{V(x+h)} \gamma(\vec{r} - \vec{s})\, d\vec{r}\, d\vec{s} \tag{2.6}$$

For a special case when the lag distance, h, is much larger than the larger of $\sqrt[n]{v}$ and $\sqrt[n]{V}$, where n is 1, 2, or 3 dimensions, then the regularized variogram approximation holds ($\gamma(h)$ is the point variogram):

$$\gamma_V(h) \approx \gamma(h) - \bar{\gamma}[V(x), V(x)] \; for \; h \gg \sqrt[n]{V} \tag{2.7}$$

V is a sample support size, which is regarded as a length, area or volume depending on the number of dimensions. The integral range, A, is the size of the domain for which we can hope to achieve ergodicity in practice. It represents a length, a surface or a volume depending on whether the problem is in one, two or three dimensions [Matheron, 1989]. The definition used in this dissertation is from [Matheron, 1989], and mathematically becomes:

$$A = \frac{1}{C(0)^2} \int_S C(\vec{r})\, d\vec{r} \tag{2.8}$$

Where $\vec{r}$ corresponds to the dimension of the spatial region S. The quantity $C(h)$ is the covariance function and is related to the variogram $\gamma(h)$ by the relation: $C(h) = C(0) - \gamma(h)$, where $C(0)$ is the field variance. The dispersion variance, $D^2(v/V)$, is the mean value over V of the estimation variance of the regularized variable $Z_V(x)$ by the regularized variable $Z_v(y)$ of a unit v located inside V. Mathematically, it can be computed using the following formula [Journel and Huijbregts, 1978]:

$$D^2(v/V) = \bar{\gamma}[V(x), V(x)] - \bar{\gamma}[v(x), v(x)] \text{where } nv = V$$

$$\text{in case } v \ll V, \text{ then we use: } D^2(v/V) = \int_{V(x)} \sigma_e^2[V(x), v(y)]\, dy \tag{2.9}$$

One important property of the dispersion variance is Krige's Relation, named after D. G. Krige [Journel and Huijbregts, 1978] who found this relation experimentally. It is the additivity property of the dispersion variance, and can be written as:

$$D^2(v/P) = D^2(v, V) + D^2(V, P), \text{ where } v \subset V \subset P \tag{2.10}$$

The significance of the formula for $\bar{\gamma}[v(x), V(x+h)]$ should be stressed, because the regularized variograms, dispersion variance and estimation variance all depend on it in one way or another. Therefore, a great deal of analysis and development was put into the computation of this equation during the work of this project.

2.3 Experimental Variograms and Dispersion Variances

The algorithms used to compute experimental variograms in this dissertation were routines available in the GSLIB package (available for free on the Internet at *www.gslib.com*, but it is better to have [Deutsch and Journel, 1997] for a complete users manual). Specifically, the programs GAM and GAMV from GSLIB were used to compute the sample variograms for data on a regular grid and on arbitrary locations, respectively. However, in some cases when there is a regular grid, and the interest is to compute efficiently and completely the omnidirectional variogram, an algorithm proposed in [Webster, 1985] was used. The algorithm is best suited for the two dimensional case of correlated random fields that are arranged in a regular rectangular shape. The method can be summarized with the following equations:

$$\begin{aligned} S_1(p,q) &= \sum_{i=1}^{n-p}\sum_{j=1}^{m-q}(Z_{i,j} - Z_{i+p,j+q})^2 \\ S_2(p,q) &= \sum_{i=1}^{n-p}\sum_{j=q+1}^{m}(Z_{i,j} - Z_{i+p,j-q})^2 \\ \gamma(h) &= \gamma(p,q) = \frac{1}{2}\cdot\left[\frac{1}{2\cdot(n-p)\cdot(m-q)}\right]\cdot[S_1(p,q) + S_2(p,q)] \end{aligned} \tag{2.11}$$

Where: $h = \sqrt{(\Delta_n \cdot p)^2 + (\Delta_m \cdot q)^2}$

n=number of nodes in the x direction.

m=number of nodes in the y direction.

p=lag in the x direction.

q=lag in the y direction.

Δ_n=distance between nodes in the x direction.

Δ_m=distance between nodes in the y direction.

i=summation index in the x direction

j=summation index in the y direction

The method shown above is simply a reorganization of the computations of the traditional estimator, which was developed by Matheron. It is deduced using the method of moments. It assumes each pair of observations to be equally probable and is very vulnerable to outliers. If the data are not Gaussian, the estimator is not robust [Cressie, 1993]. The final equation of the traditional estimator, for one dimensional analysis, is [Cressie, 1993]:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x)_i - z(x+h)_i]^2 \tag{2.12}$$

This estimator can be generalized to two- and three-dimensions, in fact the generalization to two dimensions is given in equation (2.11). The following estimator is present in several works, in particular [Cressie, 1985] and also [Russo and Jury, 1987a]. The idea is to make a more robust estimator that takes into account, say, a log-normal distributed data set. The estimator adjusts for the bias and therefore it is supposed to be more robust to outliers. The bias adjustment factor assumes the data are normally distributed [Cressie, 1993].

$$\hat{\gamma}(h) = \frac{1}{2} \left\{ \frac{1}{N(h)} \sum_{i=1}^{N(h)} |z(x)_i - z(x+h)_i|^{1/2} \right\}^4 / \{0.457 + 0.494/N(h)\} \tag{2.13}$$

And finally the following method is supposed to be even more robust. It is also presented and analyzed by Cressie in his book "Statistics of Spatial Data" [Cressie, 1985]

$$\hat{\gamma}(h) = \frac{1}{B(h)} \left[\text{med} \left\{ |z(x)_i - z(x+h)_i|^{1/2} : i \in N(h) \right\} \right]^4 \tag{2.14}$$

where med refers to the median of the data. The reasoning behind the estimator in equation (2.13) is that if the variable z is gaussian, then $(z(x)_i - z(x+h)_i)^2$ is a variable with a χ^2 distribution with one degree of freedom. The traditional estimation does not keep the gaussian characteristics of the original variable z. The transformation that makes the squared difference more a gaussian variable is a $\sqrt[4]{[]}$, represented as $|z(x)_i - z(x+h)_i|^{1/2}$,

i.e., the square root of the absolute difference. The estimators in equations (2.13) and (2.14) are location estimators of $|z(x)_i - z(x+h)_i|^{1/2}$ that include a correction for bias and are robust. Notice that the last estimator does not include any sum term and is very different from the other two estimators. The value $B(h)$ is said to be a bias correction and [Cressie, 1993] says it is equal to 0.457 asymptotically as the number of sample pairs goes to ∞. When $|z(x)_i - z(x+h)_i|^{1/2}$ is skewed, the mean and median are different so bias is present unless symmetry is present. To compute the median of a dataset, the following algorithm is recommended in [Press et al., 1992]:

1. Sort the data, x_i, in an ascending or descending order.

2. median $= x_{(N+1)/2}$ for N odd.

3. median $= \frac{1}{2}(x_{N/2} + x_{N/2+1})$ for N even.

The space requirements in terms of memory is of course much larger for estimator in equation (2.14) because the list of all the possible pairs that are separated at a said lag distance must be kept and then sorted (using, for example the QuickSort routine in [Press et al., 1992], or a more traditional Bubble Sort technique), which increases the amount of computational requirements while the estimators in equations (2.12) and (2.13) require the computation of a summation. However, the estimator in equation (2.14) has two desirable characteristics: robustness and quantile capture. [Cressie, 1993] tested the estimators in equations (2.12) and (2.13) above without much difficulty, and found out that the estimator in equation (2.13) is biased (the sill value is lower compared to the value obtained with the traditional estimator). The speed of computations was also slower for the estimator in equation (2.13) because of the higher degree of complexity involved. According to [Cressie, 1993] the estimator in equation (2.13) is more efficient in benchmark tests against the estimator in equation (2.14). In addition, there is not a formula or an explained criteria on how to choose $B(h)$. What value should one assume? maybe the asymptotic value? It is not clear how to get a reasonable estimator for that value so that the bias is reduced. As was said before, Cressie has done a

remarkable work in comparing and analyzing several methods to compute the experimental variogram. For instance, [Cressie and Hawkins, 1980] present an assortment of methods that are uncommon in the literature. Several methods proposed by Cressie are of the class of M-estimators, which imply an iterative procedure to solve at each lag (very computationally intensive). [Cressie and Hawkins, 1980] is merely a presentation and some explanation in detail, but by no means is the end.

Other authors have taken other approaches to try to reduce noise in the estimates of the $\gamma(h)$ values. For example, the work done in [Delay and de Marsily, 1994] presents an alternative idea of using the integral of the semivariogram instead of the semivariogram function directly, which can be useful when the experimental semivariogram has a high degree of variability. This reference presents an analysis using several variogram models, in particular the linear, power, spherical, cubic, exponential, and gaussian models. The gaussian model turns out to be quite difficult because one must calculate an integral of the form: $\int_0^D e^{-\left(\frac{h}{a}\right)^2} dh$, although the result can be written as the error function distribution [Abramowitz and Stegun, 1970]. A probability function, which can be easily computed (using for example [Press et al., 1992]), is used instead of an error function. The method is useful but does not give much information for the case of regularized variables and variograms. However, in the work of this dissertation the use of the method explained in [Delay and de Marsily, 1994] was not used. The applicability is beyond the scope of the deconvolution of regularized variograms, and so the idea explained in the paper does not directly apply. However, in the future a similar approach might be of interest.

2.4 Numerical Approximation of Regularized Variograms and Dispersion Variances

The multiple integrals describing the function $\bar{\gamma}[v(x), V(x+h)]$ cannot be easily calculated explicitly. The approach followed in this dissertation was to first reduce the number of multiple integrals using the *Cauchy Algorithm*. This algorithm was employed in [Ellsworth and Boast, 1996] for $\bar{\gamma}[v(x), V(x+h)]$ and is described in detail by [Journel and Huijbregts, The approach consists of using a geometric variogram defined by index functions so that the integration limits can be extended to $\pm\infty$. The details of the method are presented in Appendix A for one, two, and three dimensions. At the end, the algorithm allows the simplification of the integral into a summation of several integrals whose integration limits do not depend on the lag separation. The integrals, as presented in their final form in Appendix A, are all made homogenous in the sense that their integration limits are set between 0 and 1. This is true for one, two and three dimensions. The final expressions for the integrals are a single, double, and a triple integral for one, two, and three dimensions, respectively. The solution of the final expression was computed numerically using *Gaussian Quadrature* [Abramowitz and Stegun, 1970]. In this procedure, the function inside the integral (the integrand) is evaluated at selected points inside the interval defined by the upper and lower integration limits, $[0, 1]$ in this case. There are several versions and ways to define Gaussian Quadrature rules, depending on the integration limits, but for obvious reasons the method selected in this case was defined for integrals between 0 and 1. In the case of multiple integrals, the method was applied recursively, i.e., a Gaussian Quadrature (of a Gaussian Quadrature (of a Gaussian Quadrature)), if applicable. The number of sampled points was 16 in this case. In the book [Journel and Huijbregts, 1978] the authors recommend and suggest that 4 integration points in the Gaussian Quadrature are enough, but with the easy availability of enough computer power these days the extra burden included by selecting more integration points is not noticeable. The final implementation of the function was done

in C++. As a reference, I used [Schildt, 1996] which presents clearly the several features that make C++ a powerful and popular language. In particular, the functions that implement the one, two, and three dimensional $\bar{\gamma}$ were coded using a feature called *function or method overloading*, which basically allows for the programmer to declare several functions with the same name, as long as the returning value and/or method's signature (i.e., the input parameters of the method) are different among the functions. The block of functions implementing the numerical integrations became part of the more general variogram fitting program called GEOLAST, used in this dissertation.

Chapter 3

Variability Structure Estimation

3.1 Explicit Relationship of γ_v, $\bar{\gamma}$, and γ

This dissertation presents a method to do the deconvolution of regularized variograms. For this purpose, it is very important to make clearly present the relationship among γ_v, $\bar{\gamma}$, and γ. The explicit relationship between γ and γ_v and $\bar{\gamma}$ comes from the definition of $\bar{\gamma}$ itself. Recalling the definition from Section 2.2:

$$\bar{\gamma}[v(x), V(x+h)] = \frac{1}{vV} \int_{v(x)} \int_{V(x+h)} \gamma(\vec{r} - \vec{s})\, d\vec{r}\, d\vec{s} \tag{3.1}$$

This equation clearly relates γ and $\bar{\gamma}$ through an integral operation (convolution). But what about the other quantity, the regularized variogram γ_v? What is its relationship? According to [Journel and Huijbregts, 1978], the relationship is straightforward. For a *mixed support variogram*, i.e. a variogram based on a second order stationarity assumption whose separation vector has one end pointing to one sample support size (v) and the other end pointing to another sample support size (V), the mathematical relationship should be then:

$$\gamma_{vV}(h) = \bar{\gamma}[v(x), V(x+h)] - \frac{1}{2}\left(\bar{\gamma}[v(x), v(x)] + \bar{\gamma}[V(x), V(x)]\right) \tag{3.2}$$

A special case is when both support sizes v and V are equal. In such case, the equation is simplified to:

$$\gamma_v(h) = \bar{\gamma}[v(x), v(x+h)] - \bar{\gamma}[v(x), v(x)] \tag{3.3}$$

The other quantity of interest, but more for a formality for completeness of the presentation, is the relationship between $\bar{\gamma}$ and $D^2(v,V)$, the dispersion variance. The relationship is similar to the equation above and is the following:

$$D^2(v,V) = \bar{\gamma}[V(x),V(x)] - \bar{\gamma}[v(x),v(x)] \tag{3.4}$$

Clearly then all the quantities of interest depend on $\bar{\gamma}$, and thus depend on γ (point variogram) themselves. This is the most important aspect to keep into account for the purposes of the fitting routine of point variograms that uses simultaneously observations of point variograms, regularized variograms (single and mixed sample support), and dispersion variances. The computer program developed for this dissertation, GEOLAST, uses this remarkable property of $\bar{\gamma}$ begin the core function to be used in the data deconvolution process. The next issue to discuss is the error minimization problem needed for the deconvolution process, which is presented in the next section.

3.2 Error Minimization Problem for Variogram Deconvolution

An important part of the variogram deconvolution process is the correct formulation of the error minimization problem, which is basically a parameter estimation problem. The optimization problem for the parameter estimation of the variogram deconvolution process, used in this thesis, can be stated mathematically in this way:

$$Min\ E = \sum_{j=1}^{k} \left\{ F_i^{obs}(\gamma,h,v,V) - F_i^{com}(\gamma,h,v,V;\theta) \right\}^2 \tag{3.5}$$

This definition is equivalent to ordinary least squares (o.l.s.) as described in [Cressie, 1993]. The quantity θ represents a vector with the parameters to be estimated, i.e., the sill and correlation range or ranges of the variograms to be analyzed. This formulation was the one that was first implemented in the program and after several attemtps to modify it, I decided to come back to equation (3.20). However, the following discussion presents possible ways to

improve this formulation, but most of the results available have the limitation of considering only non-regularized variograms, and also no consideration is given to dispersion variances.

[Cressie, 1993] presents methods to fit *point* variograms and not regularized variograms. Nevertheless the analysis applies well to the problem at hand here. The quantity F is a function that depends on the point variogram and the sample volumes (for the case of mixed support variograms or dispersion variances), and so is a generalization of the different types of information that can be incorporated in the fitting routine. The o.l.s. approach can be readily modified to a weighted least squares (w.l.s.) formulation. In w.l.s., there are many ways to weigh the terms in the summation. One example is given by [Cressie, 1993] who proposes a way to assign weights to each of the differences. His example, though, is for the case of point variogram fitting only. He accomplishes this by using the following expression:

$$Min\ E = \sum_{j=1}^{k} \frac{|N(h_j)|}{\{\gamma^{com}(h_j;\theta)\}^2} \left\{\gamma^{obs}(h_j) - \gamma^{com}(h_j;\theta)\right\}^2 \tag{3.6}$$

The weighting factor makes sense since it weighs values near the origin much more, which is the most important feature of the variogram for estimation as the number of pairs available at each lag distance is largest for medium ranges of lag distances, so the ratio compensates for an imbalance with respect to larger lag distances.

But what happens when we have several types of data such as point variograms, regularized variograms, mixed support variograms and dispersion variances? How does one weigh these different data types in the parameter estimation? There does not seem to be a simple answer. Analyzing dispersion variances, for instance, does not consider any measurement of lag separation. Regularized variograms have already included several points in each of the field data. What can be a possible solution? The following formulation was an attempt to implement weighting factors for each data point used in the variogram deconvolution process of this dissertation. If we have m groups of data, and say each type of data has a N_w number of points, then a possible way to do a weighted least squares fit is using the following

equation:

$$Min\ E = \sum_{w=1}^{m} \left\{ \sum_{j=1}^{N_w} \frac{P_{wj}}{\{F^{com}(\gamma,h,v,V)_i\}^2} \left\{ F^{obs}(\gamma,h,v,V)_j - F^{com}(\gamma,h,v,V;\theta)_j \right\}^2 \right\} \quad (3.7)$$

The quantity P_{wj} is a number that resembles the number of pairs at a given lag separation h in the formula by [Cressie, 1993]. If we choose P_{wj} to be the number of data values for data type w, then this might give a heavier weight to types of data with more points in them, leaving quantities such as dispersion variances in a very unbalanced situation with respect to the other type of information. Dispersion variance data files usually have very few points compared to, for example, regularized variograms. Therefore, one might for now set $P_j = 1$ and use this criterion in the fitting routine. This is at least an improvement over an unweighted least squares routine because it makes the criterion to yield similar values for all data types. When I tried to implement this idea, there was the question of how to choose an appropriate value for P_{wj}. The issue in question is more complicated than it seems at first; the program developed for variogram deconvolution allows the simultaneous use of different kind of data and thus the weighting factor has to be defined in a very consistent way. The attempts that I made to provide with a satisfactory value of P_{wj} were not fruitful. Possibly an entire dissertation could be written on this particular subject of weighting least squares for the variogram deconvolution process. For this reason, in this dissertation the final version of the program GEOLAST implements equation (3.20) rather than equation (3.22).

[Cressie, 1985] presents some interesting points regarding weighted least squares (w.l.s.) and generalized least squares (g.l.s.). The implementation of the code to do the fitting procedure uses the Levenberg-Marquardt algorithm, as presented in [Press et al., 1992]. However, the first versions used an already available code written in Fortran, for which the basis was the work presented in [Parker and van Genuchten, 1984]. The code was originally developed to solve the problem of determining transport parameters from laboratory and field tracer experiments. The code was then modified so the model function (in the original case this was the advective-dispersive equation solution) was changed to be a regularized variogram using numerical integration. The first version I used by [Parker and van Genuchten, 1984] solved

only for one single regularized variogram (not even mixed support variograms or dispersion variances). Later on, I modified the code to simultaneously use regularized variograms, mixed support variograms, and dispersion variances as desired. Another reference that offered some interesting extra algorithms useful in the implementation of the code used for the variogram parameter estimation problem was the one presented in [Barry et al., 1988]. This computer program was the first one (after several modifications) to fit regularized variograms and dispersion variances to get the point variogram. The reference is complete and clear, althought it is a little bit outdated because it uses an old version of Fortran. The program implements the Levenberg-Marquardt method of data fitting, and it produces confidence intervals of the estimated parameters. The method used for the confidence intervals is the same expression as the one presented in [Bardsley et al., 1985], where one needs to compute the value of an F-statistic. The F-statistic can be computed with the method presented in this reference, or by using [Abramowitz and Stegun, 1970] which leads to the same results. I later changed the code into C++ completely, using the version of the Levenberg-Marquardt available in [Press et al., 1992]. Another possibility that was explored was the idea of using the jackknife procedure to estimate the variogram parameters. The method was not pursued in more detail apart from curiosity, given that the original computer code that was used implemented non-linear least squares.

As was seen, the estimation of variogram models parameters was implemented using simultaneously rather heterogeneous types of data, such as: point variogram, regularized variogram, and dispersion variance information, simultaneously. The Levenberg-Marquardt algorithm was used, because of the availability of existing working code and because of the well referenced characteristics in parameter estimation problems. There are many other possible error minimization algorithms, but in any case, intensive numerical integration must be performed to guarantee good results, as was implemented in this thesis. The next step that I considered was to incorporate nested structures, and in particular nugget effect. These issues are discussed in the next section.

3.3 Nested Structures for Nugget Analysis

Real world spatially correlated data collected in a field can rarely be described with a single permissible variogram model. Rather we often require a linear composition or "nesting" of several permissible variogram models to describe the observed variability structure. In this dissertation, I analyzed the CEC conditional simulation results with nugget and without a nugget. A possible way to deconvolute regularized variograms with the presence of a nugget variance is by the use of nested structures. Mathematically, this situation can be expressed as follows:

$$\gamma_{total}(h) = \sum_{i=1}^{K} \gamma_i(h) \tag{3.8}$$

This equation characterizes the spatial structure (total variogram) as a set made up of K distinguisable sub-spatial structures that can be of different theoretical models themselves. One example of a nested structure is the representation of a nugget effect and large scale trends. The nugget effect is described by [Goovaerts, 1997] as the nonzero discontinuity that the variogram function $\gamma(h)$ might have when h tends to zero. It relates to the measurement error and/or spatial sources of variation at distances smaller than the shortest sampling interval. So, in other words, a nugget effect may be caused by a nested structure that is not being detected by the sampling grid size used in the measuring process. Large scale trends are a different problem. In theory, for a second-order stationary field, a variogram function should converge to an asymptotic value as h increases indefinitely, getting to a plateau which corresponds to the sill of the field. But in real field data this is often not the case; the function might keep increasing instead of asymptotically approaching the sill value. Such situation may then be due to uncertainty in the observed $\gamma(h)$ or to a large scale trend. However, in this dissertation I did not consider large scale trends in the data. Nevertheless, it is important to mention that the computer program developed for this dissertation (GEOLAST, for deconvoluting regularized variograms and dispersion variances), accepts two nested structures (which, by the way, can be easily extended to many more

variograms) so that the representation of a nugget variance and/or a large scale trend can be easily done. This computer program was tested with known results from the literature, and that is the purpose of Chapter 4.

Chapter 4

Applications of Variogram Deconvolution

4.1 Field Measured Data: Cores + Donuts

The first example uses field data from a field experiment done in a $2m$ by $2m$ plot in Riverside, CA. The data were analyzed originally in two papers: [Ellsworth and Boast, 1996] and [Ellsworth et al., 1996]. The experiment consisted of three solutes applied in series under steady, unsaturated flow. The chemicals were chloride, nitrate, and bromide, applied in that order. The soil was sampled after 27 days, using a destructive sampling technique. 81 "cores" and 81 "donuts" were sampled in 20 layers, in increments of $0.1m$ down to $2m$. So, in total 3240 measurements were made. The donuts were $22.2cm$ by $22.2cm$, and the cores were $7.4cm$ by $7.4cm$ (i.e., the cores are $\frac{1}{9}$ of the area of the whole donut and core). Figure 4.1 shows the sampling arrangement. The type of soil was *Pachappa* fine sandy loam. Several quantities were measured, but for the purpose of this example, only the vertically integrated mass recovery of chloride was used. To compute the values of mass recovery, the following equation was used [Ellsworth and Boast, 1996]:

$$M_o(X_o) = \frac{1}{a_o} \int\int_{a(X_o)} \int_0^\infty \theta C(\vec{x}, z)\, dz\, d\vec{x} \tag{4.1}$$

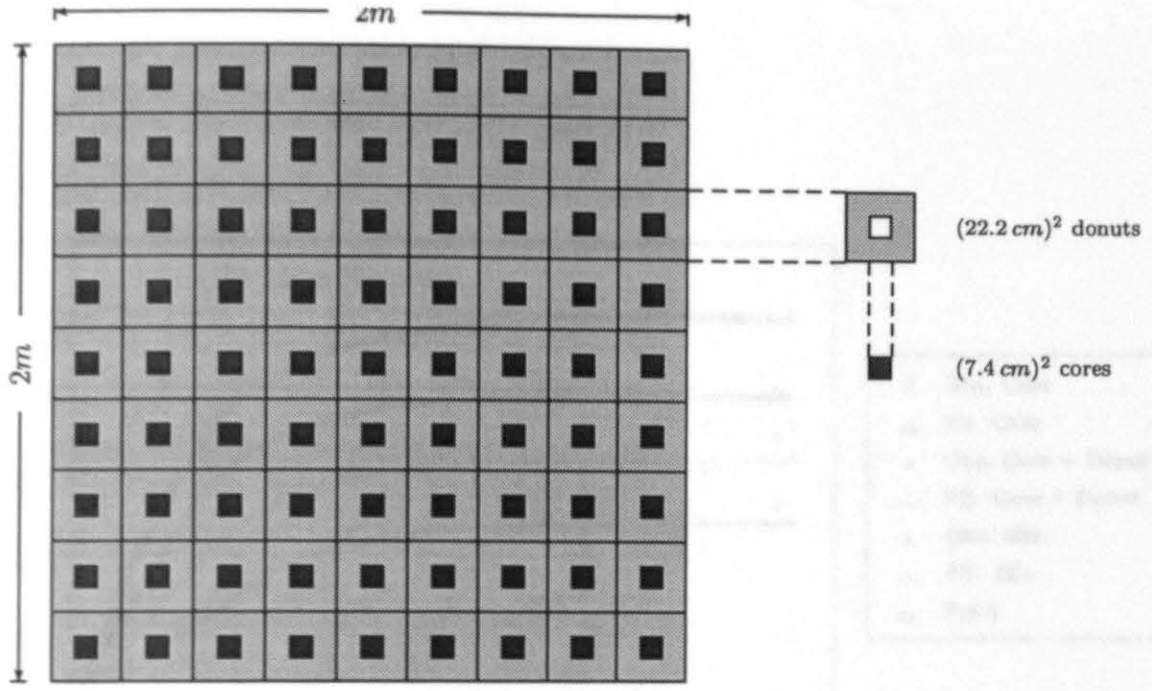


Figure 4.1: Sampling layout of "cores" and "donuts"

With the chloride mass recovery data we computed the observed variograms using the core data, core plus donut data, and then a mixed support variogram using simultaneously core data, and, core plus donut data. Also, the experimental dispersion variance was computed using the observed data. All the information was used in the data fitting process to produce an estimate of the point variogram of the field. Figure 4.2 shows the computed point variogram, as well as the core variogram, core plus donut variogram, and the mixed support case. Figure 4.3 shows the dispersion variance with observed and fitted values. The final point variogram model obtained for this equation in [Ellsworth and Boast, 1996] is $\gamma(h) = 0.11 \cdot [1 - \exp(-h/12.5)]$. The results obtained in this dissertation were: $\gamma(h) = 0.1012 \cdot [1 - \exp(-h/16.425)]$. There are some possible explanations for the differences between the two results, but basically the difference in the computation was the simultaneous use of mixed support, regularized variograms, and observed dispersion variance data. The same observations were used in both cases, but apparently the simultaneous use of the available information, including the mixed support variogram, had a different fit, in particular

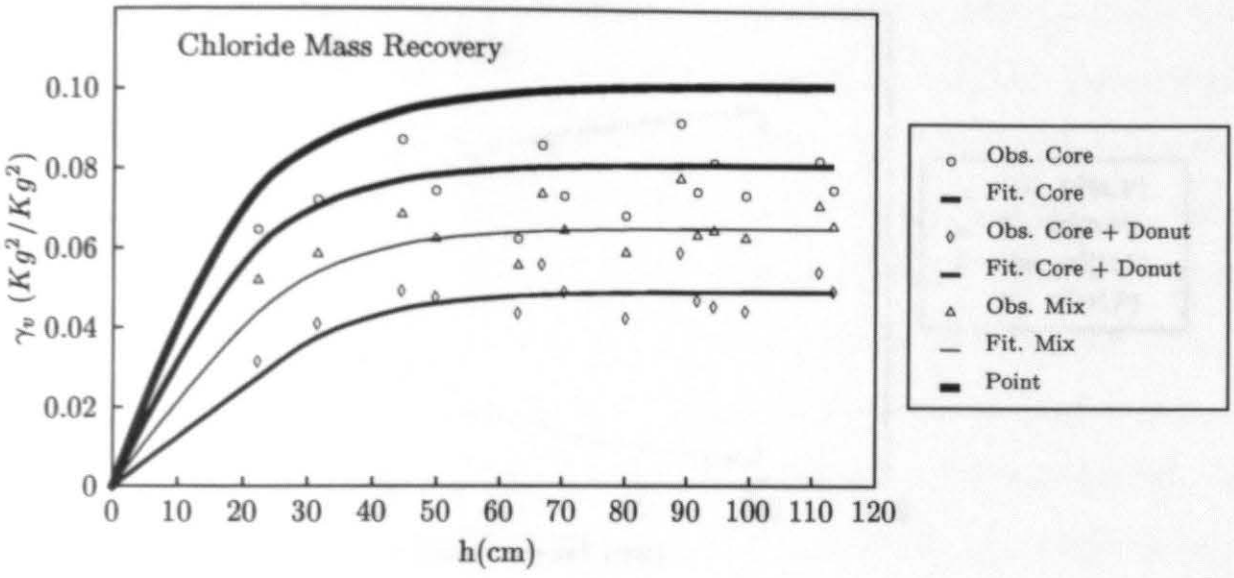


Figure 4.2: Regularized and point variograms for "cores" and "donuts"

of the theoretical dispersion variance computations. The issue of uniqueness is very important. Several runs were made using exponential, gaussian, and spherical models. No nested structures were considered in this case. The entire set of results analyzed is extensive, and for comparison purposes only the final function is presented.

4.2 Conditional Simulations

The purpose of conditional simulations is to approximate reality, keeping the same statistical features as the observed data used to do the conditioning. The approach is presented in good detail in [Goovaerts, 1997] and also in GSLIB [Deutsch and Journel, 1997]. The literature has a large number of references that show the use of the approach. For instance, in [Clifton and Neuman, 1982], the authors use three different levels of conditioning for data

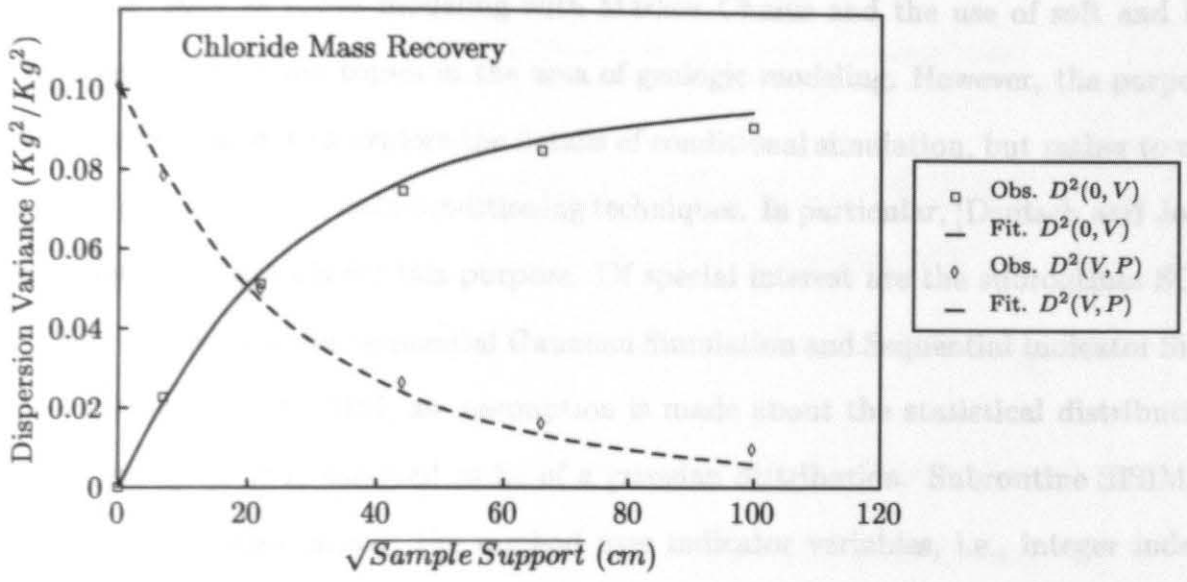


Figure 4.3: Observed and computed dispersion variances for "cores" and "donuts"

simulation. This was a notably new idea, considering the year of publication of the paper (1982). The three levels of conditioning are the following: The first one (which gives the highest level of uncertainty) happens when the log transmissivity estimates are based on measured values of the parameter but disregarding the actual spatial location of each measured point. The values of the hydraulic head computed using a multivariate normal random number generator coupled with a finite element solver have a fairly large variance. The second method consists of conditioning the log transmissivity values on existing measurements, but considering their relative spatial location, by means of kriging. The head variance is reduced to about one third of the original variance. The third method concerns the conditioning of not only the log transmissivities but also data such as flow rates and water levels in observation wells, using an inverse statistical procedure. The head prediction variance obtained with this method is about fourteen times lower than the corresponding

variance based on the kriged log estimates.

The general approach of data conditioned to existing measurements has applications not only in the generation of random fields. More advanced uses and techniques for conditioning, such as facies modeling with Markov Chains and the use of soft and hard data [Goovaerts, 1997], are topics in the area of geologic modeling. However, the purpose of this dissertation was not to explore the details of conditional simulation, but rather to use the results of proved reliable data conditioning techniques. In particular, [Deutsch and Journel, 1997] provides a set of tools for this purpose. Of special interest are the subroutines SGSIM and SISIM, which stand for Sequential Gaussian Simulation and Sequential Indicator Simulation. In the subroutine SGSIM, an assumption is made about the statistical distribution of the measured data: it is assumed to be of a gaussian distribution. Subroutine SISIM does not make such an assumption; the method uses indicator variables, i.e., integer indexes, that define different ranges of data that basically describe the entire dataset. It can be seen as a partitioning of the dataset. The method used in this dissertation was SGSIM, with a careful consideration in the decision on how to generate the conditioned data set. The conditional simulations were run for both phosphorous and cation exchange capacity (CEC) measurements. The $1024x1024$ conditional simulation of phosphorous was computed for as discussed in the publication of [Warrick et al., 1999] and so the generation process is not be repeated here. Rather, the same procedure was applied for CEC and, in a similar way the conditional simulation was analyzed as bivariate gaussian. In both cases, the resulting field was a very fine grid of $1024x1024$ points. However, when creating conditional simulations, it is never enough to run the procedure just once; rather, several simulations are run and evaluated. Such evaluation is critical to insure the conditioned data are consistent statistically with the observations. The process of generating a data set is of such importance that Section 4.2.1 is entirely dedicated to it.

4.2.1 Deciding How to Generate the CEC $1024x1024$ Data Sets

The generation of the conditional data set was a very consuming and challenging process. The original data set with measurements of the Williams Field contained several variables

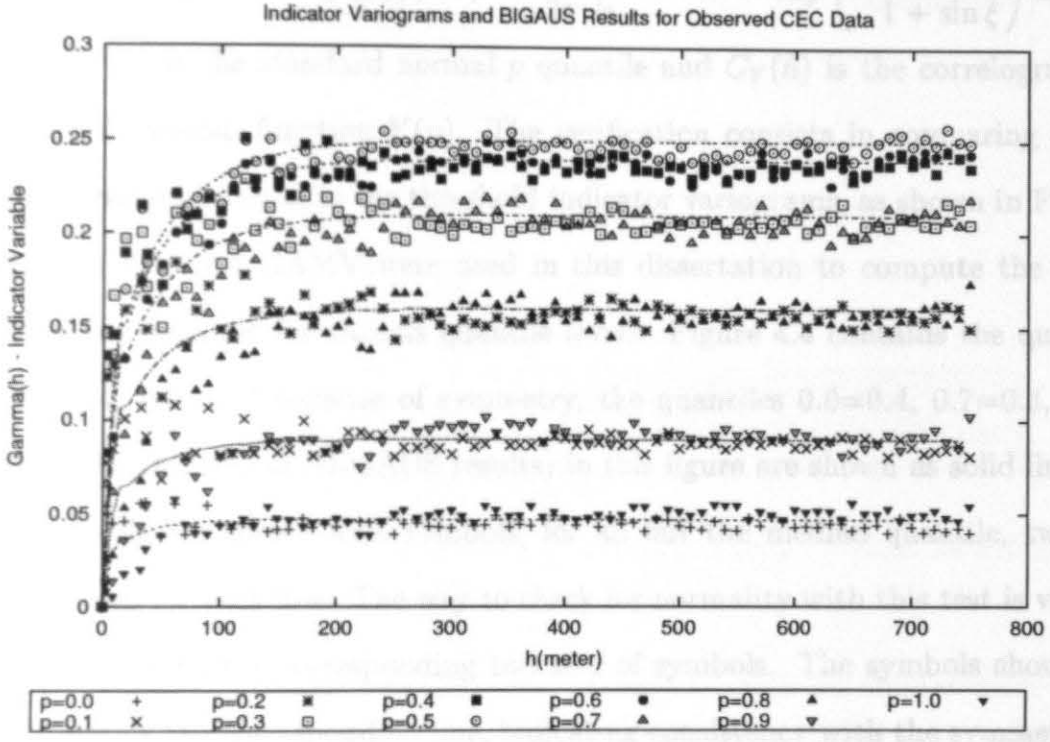


Figure 4.4: BIGAUS results compared to Indicator variograms for CEC

measured at every point. [Warrick et al., 1999] used as an example a conditional simulation for phosphorous and that example is included in this work. For the sake of studying other variables and because there were concerns about the possible existence of a trend in the phosphorous data set, it was decided to explore the possibility of using the measurements of cation exchange capacity (CEC). The idea was to perform a conditional simulation using the SGSIM program, following the guidelines in GSLIB [Deutsch and Journel, 1997] regarding the use of that subroutine. First we examine the validity of the assumption that the normal score transform of the original data set followed approximately a bivariate gaussian distribution. The normal score transform was computed using the NSCORE subroutine in the GSLIB package [Deutsch and Journel, 1997], which essentially transforms the data given

into a zero mean and a standard deviation of one gaussian distribution. I also used the program BIGAUS which is included in the GSLIB package. This program solves the bivariate probability integral [Deutsch and Journel, 1997]:

$$Prob\{Y(u) \leq y_p, Y(u+h) \leq y_p\} = p^2 + \frac{1}{2\pi} \int_0^{arcsin(C_Y(h))} exp\left(-\frac{y_p^2}{1+\sin\xi}\right) d\xi \qquad (4.2)$$

where $y_p = G^{-1}(p)$ is the standard normal p quantile and $C_Y(h)$ is the correlogram of the standard normal random function $Y(u)$. The verification consists in comparing the value given by the bivariate integral to the threshold indicator variograms, as shown in Figure 4.4. The programs GAM and GAMV were used in this dissertation to compute the indicator variograms of the data set for various quantile levels. Figure 4.4 contains the quantiles of 0.1, 0.2, 0.3, 0.4, 0.5, and because of symmetry, the quantiles 0.6=0.4, 0.7=0.3, 0.8=0.2, and 0.9=0.1. The theoretical (BIGAUS results) in this figure are shown as solid lines, while the observed levels are shown with symbols, for all but the median quantile, two sets of symbols are shown for each line. The way to check for normality with this test is visually to compare the lines with the corresponding two sets of symbols. The symbols should follow both approximately the corresponding line, indicating consistency with the symmetry of the bivariate gaussian distribution. The results in Figure 4.4 are acceptable. One idea to keep in mind is that the indicator variable to be used by the GAM or GAMV routines in GSLIB is the same whether it is defined on the original data or on the normal score data as long as the cutoff values used in GAM or GAMV are consistent with the corresponding distribution of the data. The reason behind this is that the bivariate or multivariate normality is a property of the data set and will not be affected by a linear or nonlinear monotonic increasing transform such as the normal score transform [Deutsch and Journel, 1997]. Given the agreement in Figure 4.4, I ran SGSIM routine to produce a $1024x1024$ points field. Early runs before checking for gaussianity showed a problem with data from one of the transects. The problem was that some of the points produced during the simulations were reported to come from a singular matrix in the debug file from SGSIM. The cause for this to happen is not completely known, but at least the problem was corrected by removing the information relevant to a

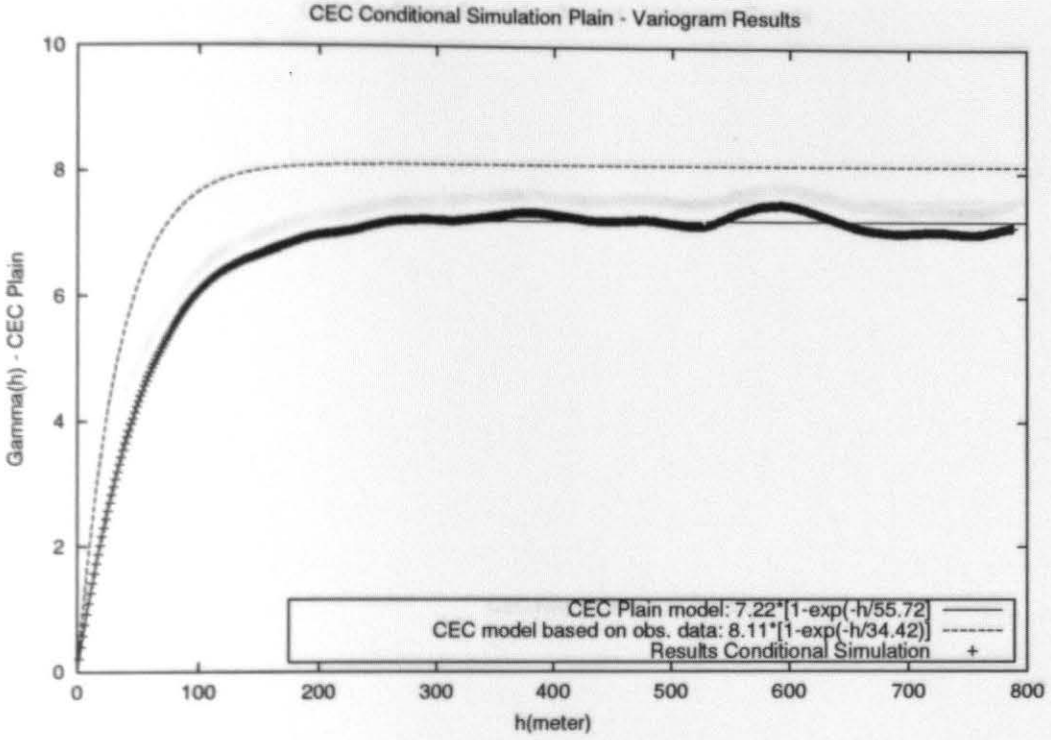


Figure 4.5: Variogram Results and Model fit for CEC Plain Conditional Simulation

specific transect (8 data points). The last part was then to compute the variograms of both the original and the simulated data for comparison purposes. Figures 4.5 and 4.6 show the resulting variograms for the CEC Plain and Nugget results respectively. The figures include the original variogram model computed from the CEC observations after the transect was removed. Figure 4.7 shows the original CEC observations variogram with the theoretical model, which was an exponential variogram model. After all the analysis performed on the CEC data, I concluded that the data set to use was the one made up of the $32x32$ and $16x16$ data together, and not including the transect information. The decision to not include the transects information at all was made by my advisor Tim Ellsworth and myself based on the observations about the statistical consistency of the conditioning data. Both cases with most of the transects (but excluding the one that caused problems before) and without any transects were analyzed. The resulting comparison of BIGAUS and indicator variograms showed a much better agreement for the no transects data file. To test for the statistical

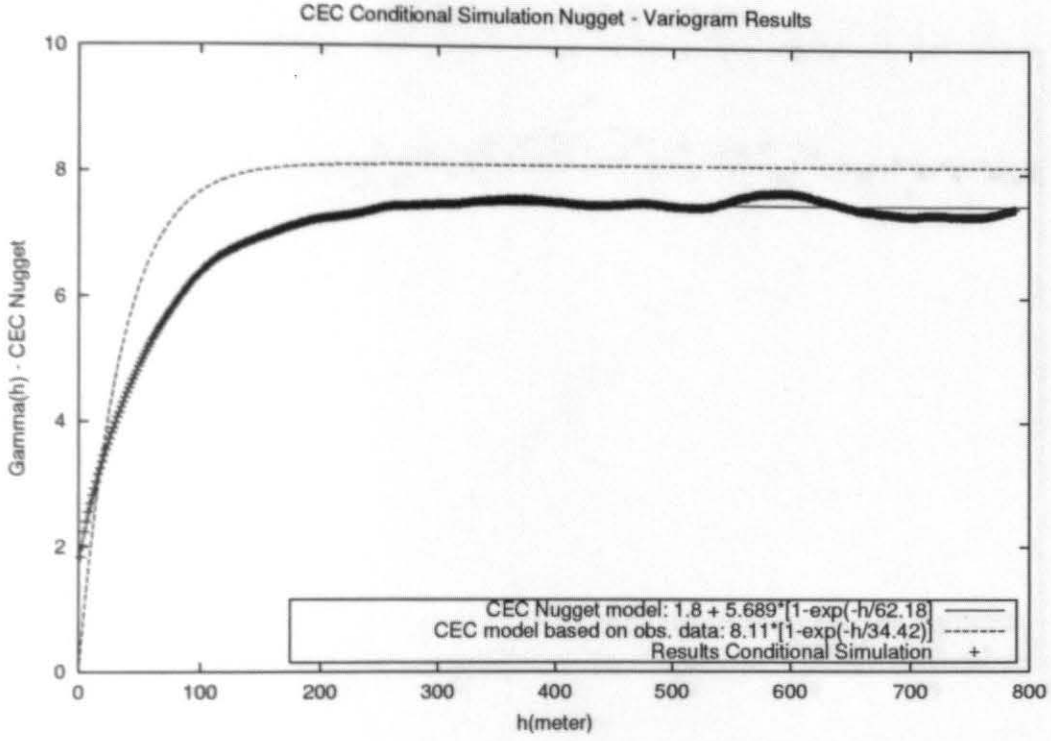


Figure 4.6: Variogram Results and Model fit for CEC Nugget Conditional Simulation

consistency of the observed data and the $1024x1024$ data set generated by the conditional simulations created, a q-q plot was used to compare the quantiles of both the original and the simulated data sets. Figures 4.8 and 4.9 show the q-q plots for the plain and nugget cases of CEC, respectively. GSLIB offers the routine QPPLT to calculate these plots in an effective manner. The idea of the plots is that the quantiles should show a nearly linear behavior. The plots presented in this dissertation show an excellent behavior of both data sets, indicating that the gaussian properties were preserved. Comparing the q-q plots for both the plain and nugget CEC cases, we can say that the q-q plot for the nugget case has a better agreement than the plain case. The reasons for this situation could be multiple, but it may be possible that the two conditional simulations are different and thus include different sets created by the random number generator. Another possibility might be the way the nugget effect is introduced by SGSIM. The original field was considered without a nugget, and so the inclusion of the nugget effect is done artificially by the program. The

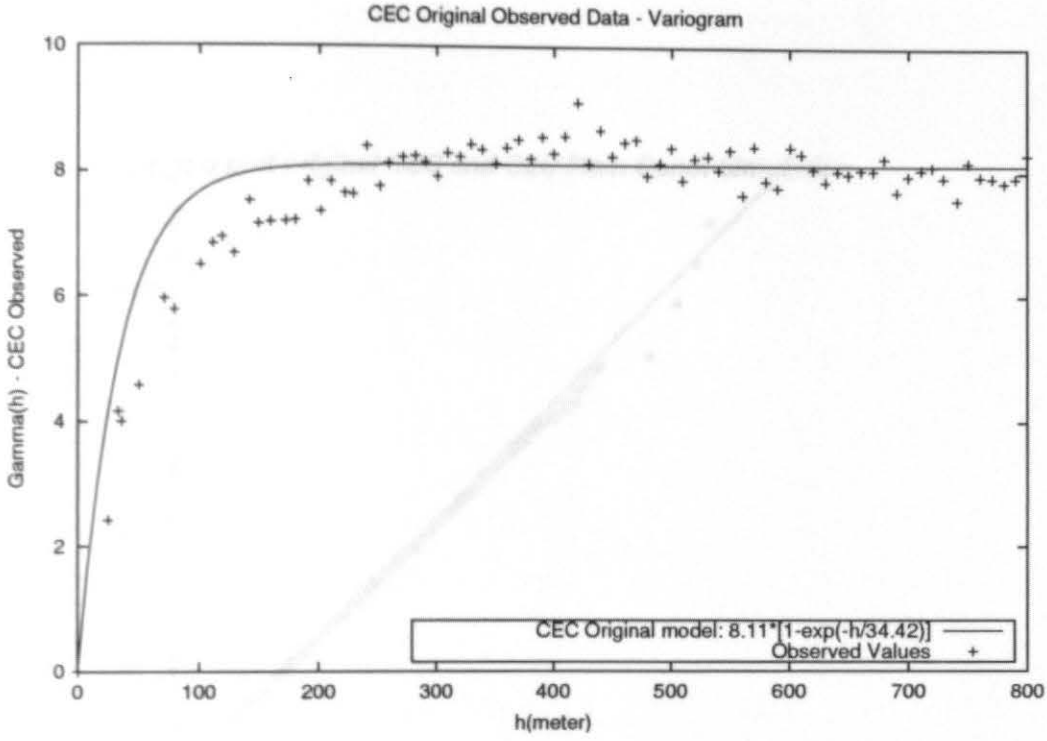


Figure 4.7: Variogram Results and Model fit for CEC Original Observed Values

exact details of the way that SGSIM does that are not clear in [Deutsch and Journel, 1997]; however, considering that artificial data drawned from a theoretical Gaussian distribution to generate a nugget effect that is not present in the conditioning data, might introduce extra points that would make the resulting conditional simulation closer to a true Gaussian distribution. In any case, the q-q plots show an acceptable agreement between the original field statistics and the conditional simulation data sets. The CEC dataset was used for the numerical experiments presented in Chapter 5. The SGSIM and SISIM subroutines allow the inclusion of a nugget effect as a percentage value of the total sill. For the CEC data, a nugget effect of 20% was included for the purposes of completeness of an analysis. The choice of this percentage of nugget effect was selected in a heuristic manner: I wanted a value of the nugget that was not too small so it would become more uncertain to determine than a larger one, but not too large so the numerical sampling experiments would not be useful at low grid resolutions (for instance, the $12x12$ cases explained in Chapter 5 would

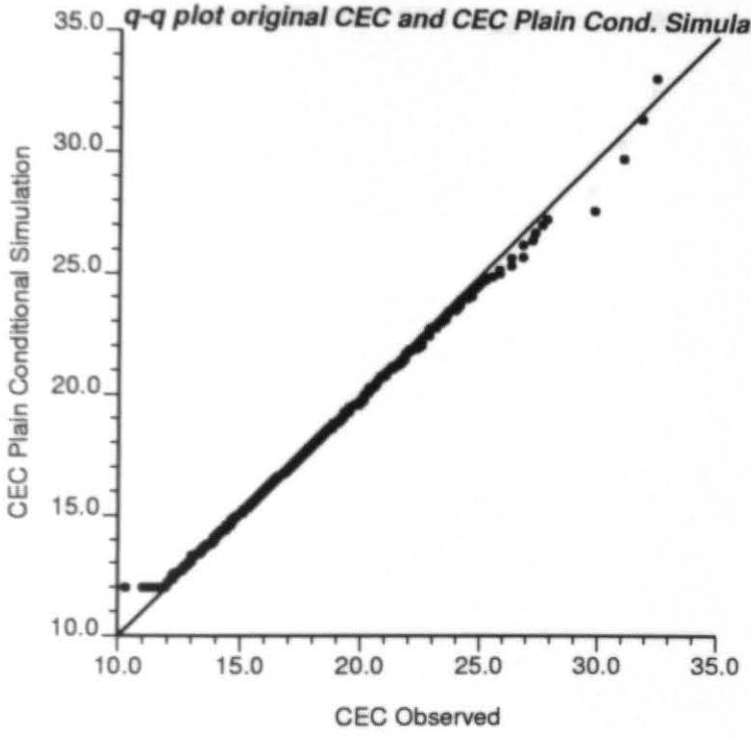


Figure 4.8: q-q plot of observed CEC values and CEC Plain Conditional Simulation Results

have had an unncessarily high uncertainty in the variogram values). So, my thesis advisor and myself decided that 20% was an acceptable number to use. One last detail about the use of the SGSIM procedure was that the value of the variance of the simulated field appeared to be quite sensitive to the choice of the seed of the random number generator. After several runs it was found that a good choice is a large number of 5 digits such as 98438 to reproduce the variance of the original field. Small numbers (i.e., five digits but < 40000) used as a seed for the random number generator would typically underestimate the value of the sill by about 30% at least. Even after this penomenon was noticed, the results used for the final CEC data sets, the largest sill I obtained was less than the sill of the original field. In any case, GSLIB always underestimates the value of the sill of the field. A possible explanation for the underestimation of the sill is that the program uses a multi-grid approach

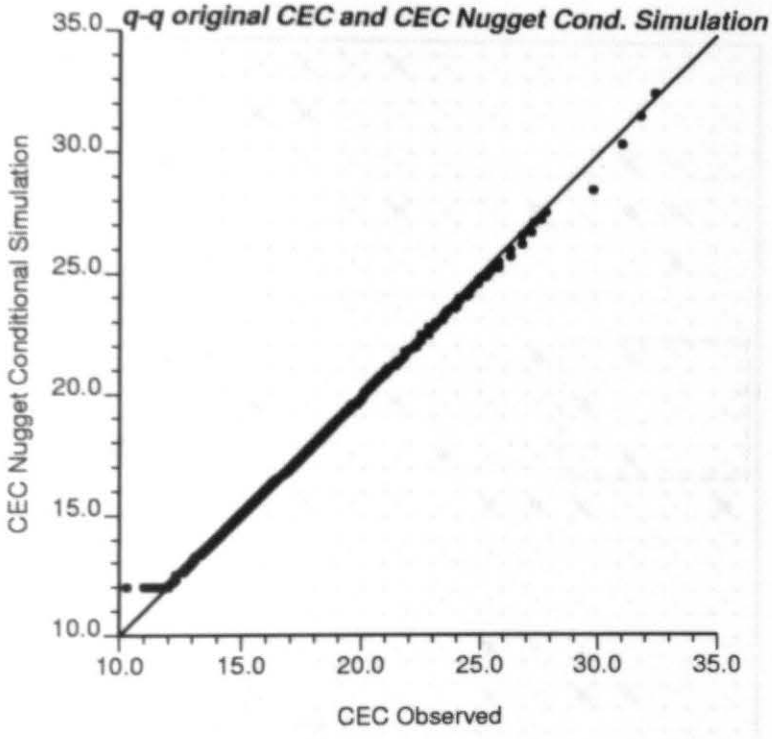


Figure 4.9: q-q plot of observed CEC values and CEC Nugget Conditional Simulation Results

to do the conditioning. The details of this method are in [Deutsch and Journel, 1997], but this approach, combined with kriging inside SGSIM, could explain a smoothing process that makes the resulting sill lower than the observed value. This is a topic that could be investigated in more detail. In any case, the value of the sill lower than the observed sill was not a major concern and so the resulting CEC conditional simulations were still acceptable for the purposes of this dissertation.

4.2.2 Williams Field: Conditional Simulation for Phosphorous

The results for this example were published before in [Warrick et al., 1999]. The conditional simulation for phosphorous was performed in a very similar manner as the one for CEC presented in the previous section. The field data for this example were taken from a field

experiment done in a 1609m by 1609m field, called the Williams field, located 50 miles NW of the city of Champaign, IL. The management is a corn-soybean rotation. The quantities

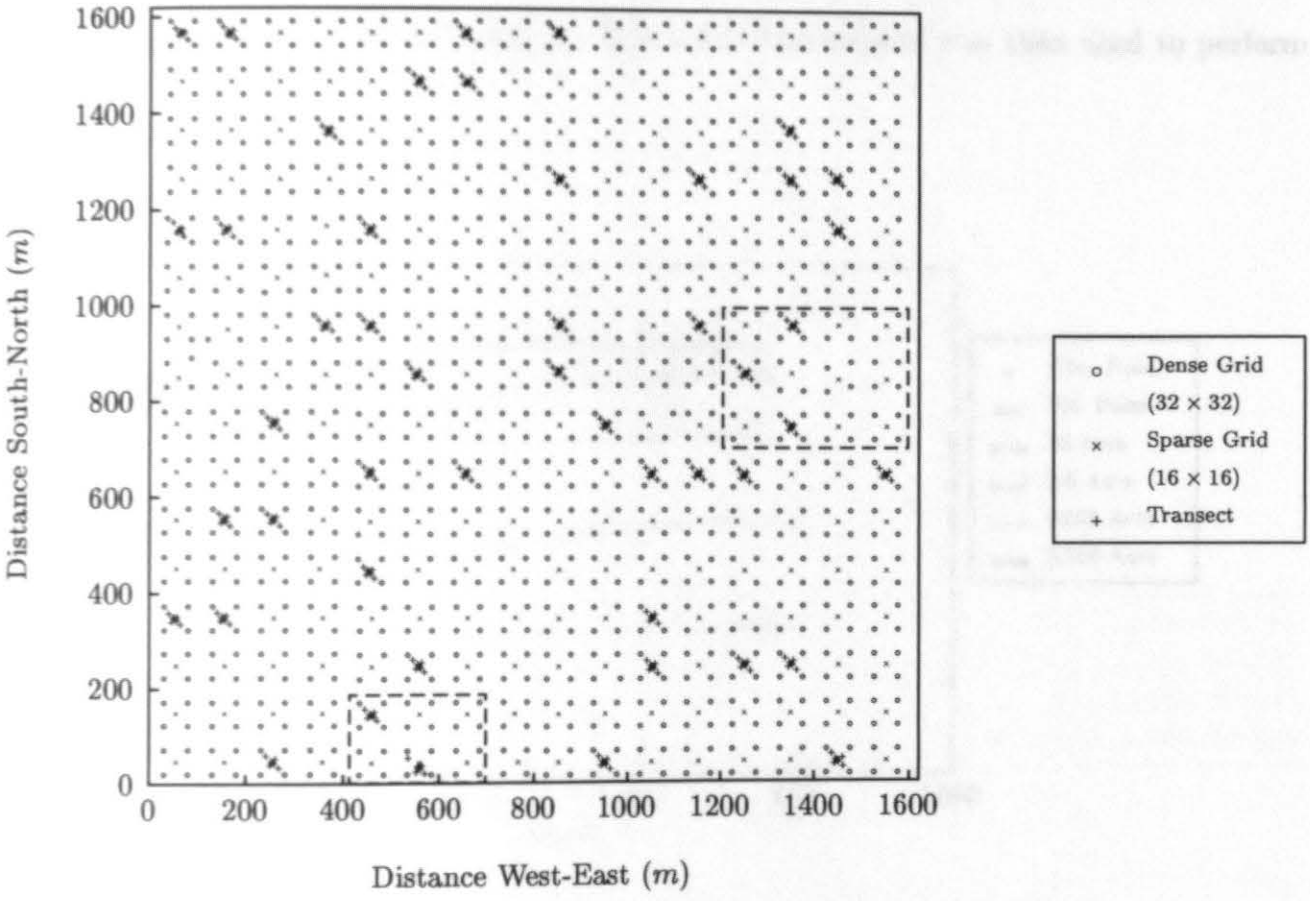


Figure 4.10: Sampling layout at Williams Field for Phosphorous Example

measured were pH, phosphorus, CEC, organic matter, potassium, calcium and magnesium. 1024 measurements were taken on a 32 by 32 grid, 256 measurements were taken at a 16 by 16 grid and 468 measurements were taken as clustered data. The details of the field soil sampling program are shown in Figure 4.10 [Warrick et al., 1999]. The areas enclosed by the dashed lines were excluded from the measurements because the data points were outside the maximum limit of the phosphorous measuring procedure (outliers). With the phosphorus data a Sequential Gaussian Simulation was performed with the field data using the program SGSIM in GSLIB [Deutsch and Journel, 1997] to get a very dense regular grid of 1024x1024

data points. The method used for conditional simulation was the same as for CEC, explained in the previous section. The separation between the points in the resulting dense grid was $1.56m$ by $1.56m$. Thus, this example dealt with a synthetic data set that was a conditional realization of a set of real observations. The dense field created was then used to perform

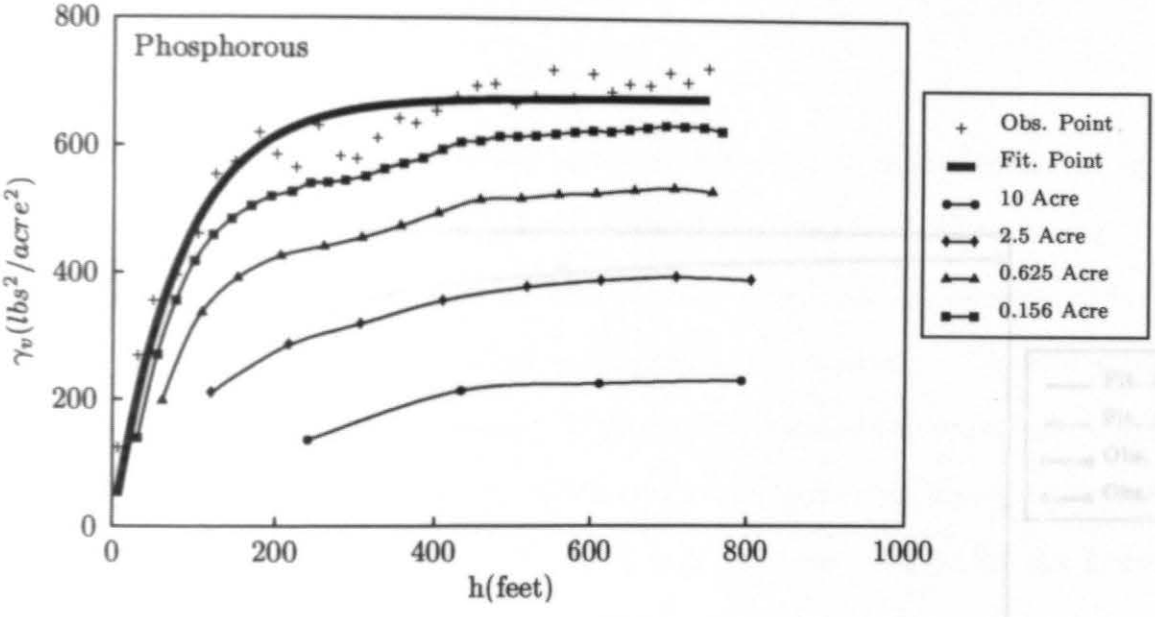


Figure 4.11: Regularized variograms and fitted point variogram for Phosphorous Example

a regularization of the data at different support volumes, i.e., clustering data points and averaging those data points in the clusters, assigning the value of the average to the center of the cluster, all of this with the help of a computer program written by the authors for the specific needs of the problem. These experimental variograms were computed for each of the regularized fields sampled this way. The experimental variograms were the regularized variograms used in the fitting routine. Also, the experimental dispersion variances were computed in a similar fasion. Using all these data, the data fitting program was used to obtain the point variogram of the field. The results of the regularized variograms are included in Figure 4.11, and the dispersion variance results are included in Figure 4.12.

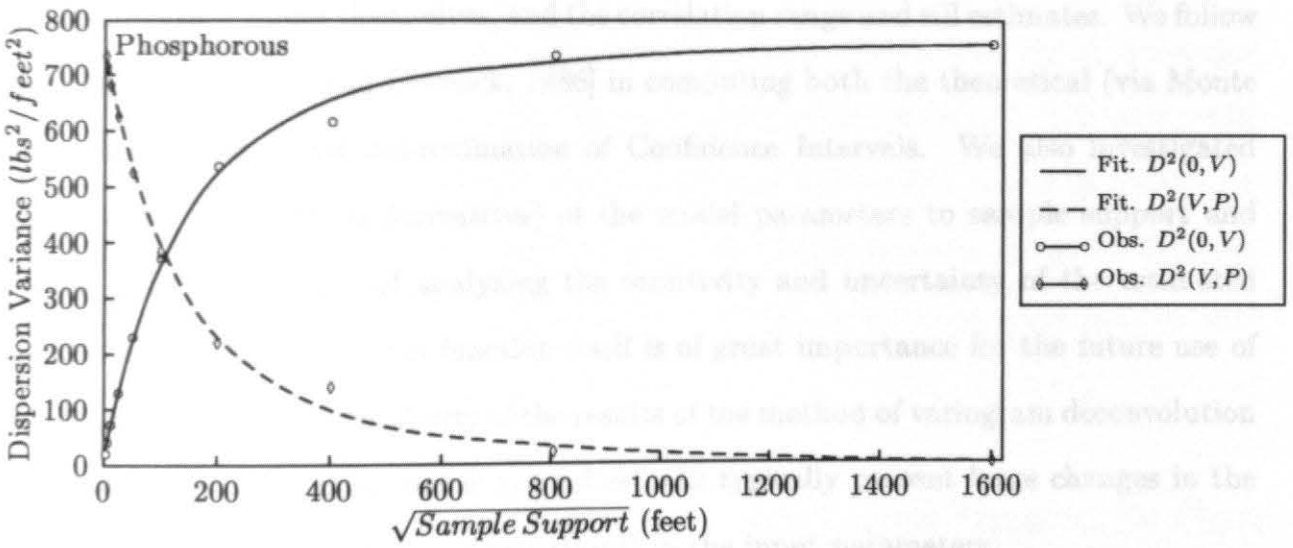


Figure 4.12: Experimental and Fitted Dispersion Variances for Phosphorous Example

Chapter 5

Robustness of the Method

We need to calculate the uniqueness and uncertainty in our estimate of the point variogram, both in the function values themselves, and the correlation range and sill estimates. We follow the approach in [Wagner and Gorelick, 1986] in computing both the theoretical (via Monte Carlo methods) and linear approximation of Confidence Intervals. We also investigated the sensitivities (i.e., partial derivatives) of the model parameters to sample support and data type. The importance of analyzing the sensitivity and uncertainty of the estimated parameters and of the variogram function itself is of great importance for the future use of any model, and also for the reliability of the results of the method of variogram deconvolution in this dissertation. A model that is not robust will typically present large changes in the results given small changes (even perturbations) in the input parameters.

Analytical solutions to the error propagation method are available for a few cases (mostly for linear models). For non-linear models, there are two approaches for applying the error propagation method: 1) Taylor method, where the forward equation is approximated by a linear function around the area of interest, and 2)the Monte Carlo approach for which one creates several realizations of the inputs (correlated random fields/transects) using a method such as SGSIM or SISIM [Deutsch and Journel, 1997]. The variability of the model outputs is then quantified by means of the variability of the output (usually confidence intervals). Monte Carlo methods do not suffer from this problem, but they do not give analytical

results which is a drawback. Monte Carlo methods were used in this thesis with excellent results. They are easily implemented and generally applicable. The contribution of error sources is assessed by means of the partitioning property: the variance of the output error is approximately equal to a sum of contributions, each of them is attribuable to an individual model input (observation, etc). In other words, $Var(out) = \sum_{i=1}^{m} Var(inp_i) \cdot (\frac{\partial F}{\partial inp_i})^2$. where F is the forward equation, which in the case of this dissertation would be the process of computing the regular variogram, where m is the number of input parameters and in the sum over $i = 1, ..., m$, i enumerate the input parameters. The partial derivative is basically the sensitivity of the forward equation to a change in input i. The analysis gives an idea of the contribution of a certain input in the model on model output and so a decision can be made about the efforts to acquire more data and in what fashion to improve the model output quality. Block kriging variance is smaller than the point kriging variance. The only really sound reason to use block kriging is the requirement that the input data be in block form (perhaps the assumption and specification of data in block form is a very attractive feature). The process of downscaling implies that variability will be added rather than removed and this is in general a difficult, if not impossible, problem.

5.1 Monte Carlo Methods for Confidence Intervals Computation

The approach used in this dissertation for using the Monte Carlo approach was to obtain 100 simulated samplings of the basic $1024x1024$ points CEC field. To accomplish this, the field was sampled first by choosing point data at nxn nodes sampling grids, where $n = 12, 16, 24, 32, 40$. This procedure was applied to both the plain field and to the field with a nugget. For each of those fields, the procedure followed was this:

1. Divide the $1024x1024$ nodes into an nxn regularly spaced grid, where $n = 12, 16, 24, 32, 40$ cells.

2. Inside each of the cells, subdivide the N-S and the E-W directions into 10 points respectively. The resulting $10x10$ sub-grid in each cell was numbered row-wise, so each of its points had a unique label.

3. A simulated sampling is then produced by systematically selecting the same point in each $10x10$ sub-grid in each of the nxn cells and making that a point in the simulated sampling.

4. This process identifies from 14,400 (for $n = 12$) to 160,000 (for $n = 40$) points. For each of these points there is, generally, a closest point in the $1024x1024$ generated field. The CEC (or other property) value at the point is assumed to equal the value generated at the closest point on the $1024x1024$ grid.

5. For each of these simulated samplings, compute the experimental variogram using the traditional formula by Matheron (method of moments).

6. Fit a theoretical exponential variogram model to each of those experimental variograms.

7. Gather the computed and the experimental variograms at each lag distance and sort the 100 computed variogram values at each lag, in decreasing order.

8. Choose the 2.5th and the 97.5th values for the 95% intervals, and the 6th and 95th values for the 90% interval. In both cases also compute the average value at each lag distance (mean estimator).

9. For each sampling strategy (i.e., 100 simulated samplings), plot the confidence intervals in the format of error bars to show the influence of sampling strategy on the confidence intervals.

10. The procedure was very similar for regularized variograms. Basically, everything is identical except that instead of picking a point inside each subgrid, a block was selected. However, this block had to be selected very carefully because a similar scheme was

desired as with the *Pachappa* field. In other words, I wanted to have a core and a composite sample as in the "cores" and "donuts" example presented in Section 4.1. The volumes (areas) of regularization used were $\frac{L}{40}$ and $\frac{L}{80}$, where $L = 1600m$, the length of the side of Williams Field. So, for both regularizations, the coordinates of the sample center were the same. In the end, a composite sample and a core sampled field are created. The computation of the mixed support variogram in this way is much more straightforward and automatic than if the two samples do not have a common center.

A similar idea was followed for the variogram parameters (sill and range) and their respective confidence intervals at 90% and 95% levels. The idea is more simplistic, though. In few steps, the procedure can be set up as this:

1. During the previous procedure, at the time of fitting the theoretical variograms to each of the 100 realizations, the program GEOLAST produces an output summary with the final variogram parameters for each case. These files are parsed with a Perl program to extract the sill and the range in each case.

2. As with the variogram values, the values of the sill and range are analyzed separately. Each group of 100 values was sorted from smallest to largest value and the same criteria to select the 95% and 90% level interval ends was used as before. Also, calculate the mean of the values to estimate the average parameter values.

The main idea of this procedure was to determine whether using one type of information or another, such as mixed support variograms or observed point variograms, is more beneficial to the computation of the point variogram parameters. The mathematical/computational tool explained in Section 3.3, that allows the simultaneous consideration of several kinds of information to estimate the parameters in the point variogram process (i.e., variogram deconvolution), was employed. The process to analyze this was to first consider each type of information by itself (point, regularized variograms, mixed support variograms) and then

to consider the effect of adding up each type of information. The graphical results, along with summary tables of the Monte Carlo analyses are presented in Appendices B, C, D, and E. The graphic plotting part was automated with Perl scripts and templates to use in the GNUplot package, which is freely available on the Internet at *www.gnu.org*. This program has been ported to all kinds of platforms, including several versions of Unix, Linux, MS-DOS, and MS Windows. The versions used in this dissertation were the Linux and MS Windows versions, assuring the same plotting format with a template for the graphics definitions. The main advantage of the GNUplot program is that it can create several kinds of output (visually and to files for postprocessing). The output format used in this dissertation was PostScript because it is easy to incorporate in LaTeX 2_ϵ, which is the tool used to write this document. The analysis was done to fields generated with and without nugget, and this was done with three numerical experiments that next will be explained and presented in more detail in Section 5.2. Also, the same procedure applied to the variogram uncertainty analysis was applied to histogram analysis. This is straightforward once different sample support data are in hand, and the next section explains how was this done.

5.2 Histogram Transformations for Change of Sample Support

The effect of sample support on the estimation of spatial variability structure also considers the analysis on the distribution of the data. This was accomplished by analyzing the effect on the histogram of the data set. The general idea is that the effect of sample support makes the histogram more narrow and more symmetric than in the original point data set. This is of course related to the variogram, in particular the sill reduction effect because the width of the histogram is related to the value of the sill, therefore if we know how much reduction there is in the value of the sill, then we will have a good estimate of how much reduction there is in the spread of the data. There are several possible ways to transform

the histogram of a point process to the one of a regularized process. [Cressie, 1993] mentions two approaches, namely *Affine Correction* and *Indirect Lognormal Correction.* The first one, Affine Correction, consists of the following formula:

$$q' = \sqrt{f} \cdot (q - m) + m \tag{5.1}$$

where m is the mean of the two distributions, q and q' are the quantiles of the distributions (original and transformed one, respectively). The meaning of the parameter f is this: If the variance of the original data set was σ^2, then the distribution of the transformed variable is $f \cdot \sigma^2$. This method assumes that there is not an increase in symmetry when the support increases, and also assumes that the mean and the skewness of the distributions remain unchanged. For data that are non-negative, such as solute concentration, soil fertility levels, etc, this approach may not be desirable, as it allows negative values in the transformation. The second method mentioned in the literature, the Indirect Lognormal Correction, is summarized with a simple equation:

$$q' = a \cdot q^b \tag{5.2}$$

where:

$$a = \frac{m}{\sqrt{f \cdot CV^2 + 1}} \left[\frac{\sqrt{CV^2 + 1}}{m} \right]^b \tag{5.3}$$

$$b = \sqrt{\frac{\ln(f \cdot CV^2 + 1)}{\ln(CV^2 + 1)}} \tag{5.4}$$

where m is the mean and CV is the coefficient of variation of the original field. The formula does not perform well for values that are not exactly log-normally distributed. Thus, the mean of the transformed field must be rescaled with the formula:

$$m' = \frac{m}{q''} q' \tag{5.5}$$

where the primed quantities refer to the data that had just been transformed before. This method lowers the skewness and increases the symmetry of the transformed field, and keeps the minimum value at zero. The methods just presented assume that one knows the value of

f which is the reduction of the variance. The methods assume that the values at the point scale are known and that we want to know the histogram of the regularized field. This is not quite the case most of the time, because as was explained before, normally the point values are unknown. Rather, the regularized values are known, we may need to estimate the variogram and the histogram of the point values. It turns out that it is possible to use a known type of information pertaining to the known field: the dispersion variance and Krige's relation. Recall that the meaning of the f factor is:

$$f = \sigma^2_{block}/\sigma^2_{point} \tag{5.6}$$

and so $f \leq 1$. When the field is subdivided into n equal blocks of area v such that $V = n \cdot v$, or when $v \ll V$, where V stands for the entire field area, then the following relations apply:

$$D^2(0,V) = \sigma^2_{point} = \bar{\gamma}(V,V) - \bar{\gamma}(0,0) \tag{5.7}$$

$$D^2(v,V) = \sigma^2_{block} = \bar{\gamma}(V,V) - \bar{\gamma}(v,v) \tag{5.8}$$

Therefore we can estimate f:

$$f = D^2(v,V)/D^2(0,V) \tag{5.9}$$

Using Krige's relation:

$$D^2(v,V) = D^2(0,V) - D^2(0,v) \tag{5.10}$$

$$f = 1 - D^2(0,v)/D^2(0,V) \tag{5.11}$$

These two quantities are easy to compute once the fitting of the variogram parameters is done and so numerical approximations of these values are straightforward. Thus, for each simulated sampling, a point variogram was estimated. This point variogram model was then used to compute f using the relations given in equations 5.36 - 5.40. The computation of the dispersion variance, for a given variogram model, was computed with a short program called DVCALC that was written in C++ and was part of the GEOLAST group of programs. This program was invoked from the Perl scripts that were written to perform the Monte Carlo

analysis of histograms. In this way, 90% and 95% confidence intervals were produced for point, $\frac{L}{80}$, and $\frac{L}{40}$ sampled information. The results for this part are presented in Appendices B and C.

5.3 Numerical Sampling Experiments: Monte Carlo applications

The purpose of sampling in a geostatistical design is to be able to catch the spatial correlation range and also larger scale trends in an efficient way, usually with a restricted budget for data sampling and analysis. In this dissertation, three numerical experiments were developed to evaluate the influence of sample design on variability structure estimation. Both experiments started with two conditionally simulated data sets of CEC in the Williams Field, computed as described in Section 4.2 of this dissertation. The size of the created field was $1024x1024$ points, in a field of $1610x1610$ meters. This $1024x1024$ field was sampled with resolutions of regular grids of nxn, where $n = 12, 16, 24, 32, 40$ grids.

The exact details of each experiment are presented in the next sections, but a general description is given below. The main difference between the experiments is how samples were chosen within each of the regular grids of nxn cells. The two $1024x1024$ simulated fields differed in that one field had a nugget (20% of the sill value) variability structure and the other did not. In the following, the sampling procedures for each of the two methods is presented, followed by an explanation and discussion of the variogram computation and fitting for each of the methods. The experimental variograms in both cases were easily computed using the program GAMV in the GSLIB package [Deutsch and Journel, 1997], which can handle the case of a mixed support variogram computed in the presence of several support volumes in the same data set. Because there were 100 simulated samplings for each sampling scheme and grid resolution, the amount of information to handle increased exponentially with each new resolution and sampling scheme considered. Therefore, to make

the process as automatic as possible, a set of Perl programs was used that took care of creating the necessary input files for GAMV. From each file of each simulated sampling at each nxn resolution, the following six experimental variograms were created:

1. Point information only.
2. $\frac{L}{80}$ information only.
3. $\frac{L}{40}$ information only.
4. mix_1, made of Point and $\frac{L}{80}$ information.
5. mix_2, made of Point and $\frac{L}{40}$ information.
6. mix_3, made of $\frac{L}{80}$ and $\frac{L}{40}$ information.

The program GAMV creates the variogram points in column format, unfortunately producing several rows of zeros, and also printing the data for the six experimental variograms sequentially in the same text file, which made it a difficult process to retrieve the neecessary data. This problem was solved once again with the help of a Perl program to automatically extract the desired information in a format suitable for the program GEOLAST. Having solved this problem, the next step was to identify an optimal point variogram model for each of the variograms of each of the simulated samplings.

In the fitting process, each of the six variograms mentioned above was considered by itself. In addition, the following cases were also considered:

1. Point variogram + $\frac{L}{80}$ variogram.
2. Point variogram + $\frac{L}{40}$ variogram.
3. Point variogram + $\frac{L}{80}$ variogram + $\frac{L}{40}$ variogram.
4. Point variogram + $\frac{L}{80}$ variogram + $\frac{L}{40}$ variogram + mix_1 variogram + mix_2 variogram + mix_3 variogram.

The purpose of the last case was to analyze the merit of including all the information computed from the sampled fields, with respect to the variogram function parameteres (sill and range), and also on the estimates of the point variogram values themselves. After all the fitting cases were computed, then I proceeded to compute the confidence intervals for the variogram parameters and the variogram estimates themselves. For future research, the use of the Perl [Schwartz and Christiansen, 1997] language is strongly recommended, not only for its capabilities as a job scheduler and input text formatting, but because of its easy portability and ease of use and similarity with the C programming language. Perl is an interpreted language, and as such maintaining the code written in it is an easier task than a compiled language. Also, the interpreter programs are available for free from the Internet for a wide variety of computational platforms. The majority of the work in this dissertation was performed on a Pentium II at 233 MHz computer, with 128 MB of RAM, running the Linux operating system (Slackware Version 3.6). Part of the post-processing results were run on the same computer but using the Windows 98 operating system. The processing speed and true multi-tasking capabilities of Linux over Windows 98 were clearly evident through all the computations.

5.3.1 Numerical Experiment Number One

In this experiment, the field was systematically sampled with the following sampling supports: points, $\frac{L}{80}$, and $\frac{L}{40}$ blocks. Each one of the nxn grid cells was further subdivided into a $10x10$ sub-grid, equally spaced in X and Y, and so giving 100 cells in each of the the sub-grids. Each of these 100 sub-grid cells were assigned a sequential label from 1 to 100 to uniquely identify the relative location of each subgrid cell within each of the larger grid-cells. Then the same sub-grid cell (i.e., from 1 to 100) was systematically chosen from each nxn grid-cell to create the simulated sampling. Note that the initial points were not selected randomly, so this might be a source of bias in the results. Next, 100 simulated samplings of each sampling resolution were created, making sure to get one set of 100 samplings for point,

another set of 100 samplings for $\frac{L}{80}$, and another set of 100 samplings for $\frac{L}{40}$. Thus, for each of the values of n, 300 files were created. The center of the blocks coincided with the point value and the centers of the $\frac{L}{80}$ and $\frac{L}{40}$ samples, analogous to the design shown in Figure 4.1 (core + donut), such that at the end the $\frac{L}{40}$ was a composite of the $\frac{L}{80}$ block. The sampling procedure was done automatically under Linux using several Perl programs to have a sound and stable book-keeping protocol to keep track of computations.

The simulated samplings generated were used to compute the experimental variograms, also with the help of Perl under Linux, and with the help of a program written in C++ that uses the algorithm mentioned in Section 2.3, making sure that the maximum lag distance was restricted to about $\frac{1}{2}$ of the length of the side of the field [Journel and Huijbregts, 1978]. Each of the experimental variograms was then fitted to an exponential model (using GEOLAST), again using programs in Perl to take care of preparing the input file and of processing correctly each of the experimental variograms, keeping all the files correctly labeled. The fitting procedure was applied to the files by themselves, i.e., the experimental variograms created from the point data alone, the $\frac{L}{80}$ data alone, and the $\frac{L}{40}$ alone. In addition, the point variogram model was also estimated using the mixed support and the case where all the information was included to study the effect of data availability on the values of the estimates.

Having done this, the next step was to create the confidence intervals of the variogram parameters and of the variogram values as well. The procedure that was employed for this was already explained in Section 5.1 of this dissertation. The results for numerical experiment number one are included at the end of this dissertaion in Appendices B and C for the no nugget (plain) and nugget cases, including all the variograms, histograms and tables summarizing the results for the point variogram parameters.

Background Information for Numerical Experiments Two and Three

Numerical experiments number two and three provided sampled data files that had 576 points with a mixture of sample supports. The difference between both methods was how the mixture and location of sample supports was selected. Also, experiment number three used a sampling scheme *with replacement* which makes it non-comparable with either experiment two or experiment one, for practical purposes. Before going into specific details, note that at the end both experiments number two and three produced a fairly random field with a mixture of sample support volumes (areas) that made the computation of the experimental variograms challenging. The variograms were once again calculated with the GAMV routine in GSLIB [Deutsch and Journel, 1997]. The amount of information generated, as well as graphs, was not as large in experiment number one. The reason for this is because experiments number two and three concentrated on the $12x12$ and $24x24$ grids as a way to add randomness in the sampling procedure. The procedures followed to compute the confidence intervals of both the variogram functions and the parameters of the resulting fitted point variograms is as outlined in Section 5.1. The resulting confidence intervals, for both the variograms and the parameters are included in appendices D and E for both plain and nugget fields.

5.3.2 Numerical Experiment Number Two

Numerical experiment number two was performed using a grid resolution of $24x24 = 576$ points. The procedure followed was to create a set of 100 simulated samplings, each of them created as follows:

1. Select the initial point systematically as in experiment one. To do this, each of the $24x24$ cells is subdivided in a $10x10$ regular sub-grid. Each of the 100 resulting points is assigned a unique relative label. For each realization, pick the point labeled with the same index as the realization number.

2. At the selected subgrid cell within each of the nxn grid-cells, assign a sample support size from three possible ones: point, $\frac{L}{80}$, or $\frac{L}{40}$. This assignment is done by drawing the values of the values 1, 2, and 3, with equal probability.

3. Compute the sample variograms for each of the support volumes, as well as the mixed variograms of: point and $\frac{L}{80}$ *(mix1)*, point and $\frac{L}{40}$ *(mix2)*, and $\frac{L}{80}$ and $\frac{L}{40}$ *(mix3)*. To do this, the GAMV program supplied in the GSLIB package [Deutsch and Journel, 1997] was used, in combination with Perl programs to automate the process.

4. Fit the sample variograms computed from the previous step, in addition to three cases: point + $\frac{L}{80}$, point + $\frac{L}{40}$, $\frac{L}{80}$ + $\frac{L}{40}$, and all the variograms at the same time.

5. Compute the confidence intervals for the point variogram itself, and also for the estimated parameters of the variogram function.

A typical realization of this procedure is shown in Figure 5.1. All the results for this experiment are presented in Appendices D and E for plain and nugget fields, respectively. The random number generator used is the one included in the official Perl program, which is a reliable random number generator drawing values from a specified interval or set (in this case the set $\{1, 2, 3\}$ and from a uniform distribution). When the procedure is finished, each sampled field of each realization will have, on average, an equal amount of each of the point, $\frac{L}{80}$, and $\frac{L}{40}$, or in other words: $\frac{576}{3} = 192$ points for each of the sample supports. However, due to the randomness, each realization is not exactly distributed like that and one of the sample support sizes will have more points compared to the other ones.

5.3.3 Numerical Experiment Number Three

Numerical experiment number three used a $12x12 = 144$ regular sampling grid as a template. However, as will be shown, the total number of points to be sampled was the same as for experiment number two. The purpose of experiment three was to evaluate the influence

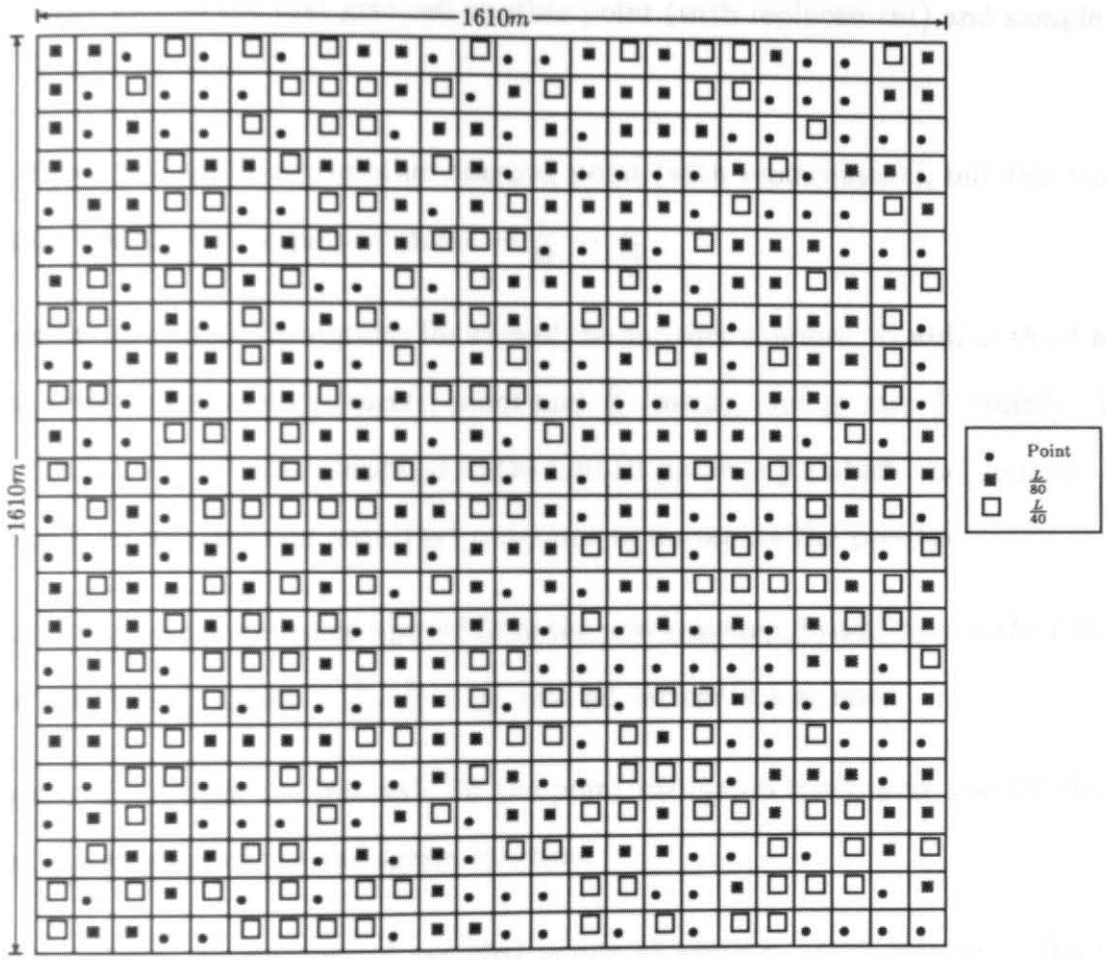


Figure 5.1: Typical realization Experiment 2: 24x24

of alternative random sampling relative to the random sampling of experiment two. The procedure followed for each of the 100 simulated samplings is as follows:

1. In the first of the 144 grid-cells, randomly pick a starting point, from a uniform distribution between 1 and 100 (with replacement) from within the grid-cell and do point sampling at that relative same location in all the grid-cells.

2. Again, in the first grid-cell, pick another point randomly from a uniform distribution between 1 and 100, (with replacement) and sample at that relative same location in

all the grid-cells with $\frac{L}{80}$.

3. Randomly pick in the first grid-cell another point (with replacement) and sample with $\frac{L}{40}$.

4. As before, pick randomly another starting point (with replacement), but this time go by rows sampling in sequence: point $\rightarrow \frac{L}{80} \rightarrow \frac{L}{40}$

5. Compute the sample variograms for each of the support volumes, as well as three mixed variograms: point and $\frac{L}{80}$ (*mix1*), point and $\frac{L}{40}$ (*mix2*), and $\frac{L}{80}$ and $\frac{L}{40}$ (*mix3*). To do this, the GAMV program supplied in the GSLIB package [Deutsch and Journel, 1997] was used, in combination with Perl programs to automate the process.

6. Fit the sample variograms computed from the previous step, in addition to the following cases: point + $\frac{L}{80}$, point + $\frac{L}{40}$, $\frac{L}{80}$ + $\frac{L}{40}$, and all variograms at once.

7. Compute the confidence intervals for the point variogram itself, and also for the estimated parameters of the variogram function.

With this procedure, the number of sampled points at every support volume is the same, and the total number of sampled points for all the support volumes is again 576. A typical realization of this procedure is shown in Figure 5.2. The results, including tables and graphs of variograms and confidence intervals, are presented in Appendices D and E. Once again, the random number generator used was the one provided in the Perl program distribution which has been proven to be reliable and accurate. This generator can select a value from a uniform distribution, as in selecting a value from a finite list of numbers such as $\{1, 2, 3\}$

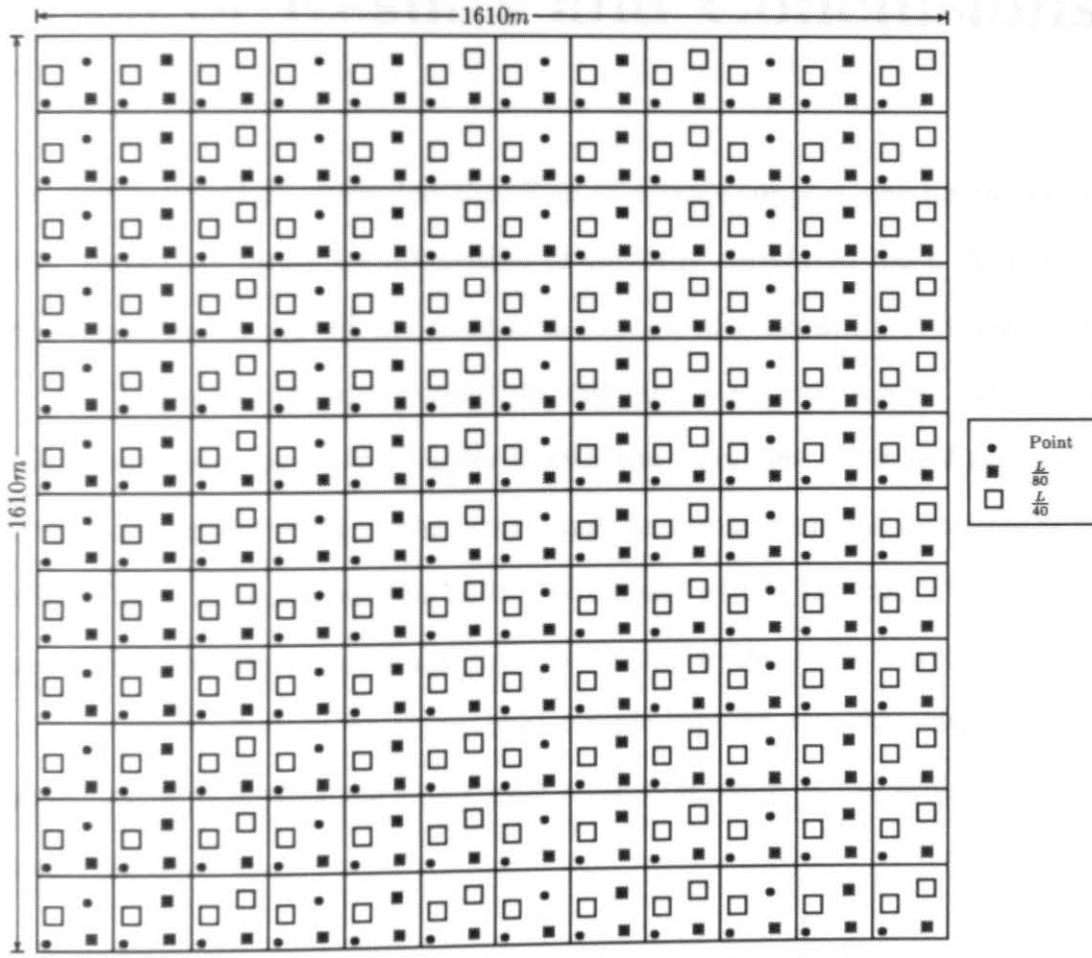


Figure 5.2: Typical realization Experiment 3: 12x12

Chapter 6

Analysis of Results and Conclusions

The work in this dissertation presented the effect of sample support size on the estimation of the spatial variability structure estimation of spatially correlated data. The work presented in this dissertation proved the method is capable of estimating a variogram that is consistent with different scales and resolutions and sample support sizes. The capabilities of the methodology were shown in Chapter 4 with two examples using real measured field data, one from a field experiment in California, and the second example a numerical realization conditioned on field measurements in the Williams Field in Illinois. The results of the examples show that the use of regularized data to estimate the point support variogram is feasible and that there seems to be an optimal sample support size that gives good results without seriously compromising the small scale resolution which is the problem with larger sample support sizes.

Comparing Experiment 1, Experiment 2, and Experiment 3

Experiment number 1 considered a systematic, non-random approach to sampling, using resolutions of nxn points, where n = 12, 16, 24, 32, and 40. Experiment number 2 used a grid of $24x24$ points but considering the selection of the sample support at each of those points to be a random variable. Experiment number 3 was a sampling exercise on a $12x12$ grid, where the initial sampling site for each of the sample supports was randomly chosen.

The total number of sampled points in experiment number 3 is exactly 576, with 192 samples at each of the sampling supports considered: point, $\frac{L}{80}$, and $\frac{L}{40}$. In experiment number 2, the number of sampled points for each sample support was $\cong$ 192 because of the randomness in the choice of sample support. The ensemble average of the number of points would be close to 192, but each realization varies in how these numbers are distributed. Regardless, the total number of points sampled per realization in experiment two was always 576. This situation makes it difficult to directly compare experiments two and three with experiment one. However, it is possible still to make comparisons related to the confidence intervals widths, and how these change with respect to increasing the number of sampled points.

Consider figures B.2 and B.18. B.2 is for the case of $12x12 = 144$, whereas B.18 is for the $24x24 = 576$ case. The number of sampled points is a 4 fold, and analyzing the confidence intervals for the variogram function, it gives a 50% reduction of the width. Now compare figures B.18 and B.34. In this case, the number of points compared is $24x24 = 576$ (B.18) versus $40x40 = 1600$ (B.39), so it is about 2.78 times. The reduction in the spread of the confidence intervals is about a half. We can conclude then that the effect of increasing the sampling grid about 11 times (from $12x12$ to $40x40$), reduces the variogram function variability to about $\frac{1}{4}$ of the original size, for experiment 1.

Figures D.1 and D.11 correspond to a similar situation for numerical experiments two and three, respectively. In this case, though, the comparison is based on the source of randomness in the sampling scheme. Recall that experiment two considers the sampling support random, while experiment three considers the initial reference point to be random. Yet when comparing both figures, we can see that the confidence interval spread is about twice for experiment number two than that for experiment number three. This tells us that a systematic sampling grid provides a better representation of the variability structure, and as randomness is introduced into the relative sample locations, the uncertainty increases. This characteristic has been discussed in [Cressie, 1993]. So, imposing a random selection of sample support does not improve the uncertainty in the variogram confidence intervals

spread. This is evident regardless of the sample support chosen. Keep in mind that a basic difference between experiments two and three is that experiment 3 still considers regular sampling grids for each one of the sample supports, whereas experiment 2 does not have regular sampling grids at all.

Turning our attention to Figures D.1 and B.2, we can compare now similar results between a $12x12 = 144$ points, and $\frac{576}{3} = 192$ points for experiments one and two, respectively. Both experiments have a comparable number of points, but the spread in the confidence intervals for experiment two results (D.1) is about twice the experiment one results (B.2). Recall that experiment one does not have any randomness while experiment two and experiment three include certain degree of randomness each. Therefore, the effect of introducing randomness in the numerical experiments actually increased the uncertainty in the estimates of the variogram function [Journel and Huijbregts, 1978]. On a similar note, from the previous paragraph comparing experiments two and three, we can conclude that as the randomness introduced increases, then the estimates of the variogram function become more uncertain.

Randomness and the use of regularized variograms

As we have already observed in the previous section, the introduction of randomness in the sampling scheme (both in the sampling sites and in the choice of sample support) introduces a higher degree of uncertainty in the variogram function values. Now the question turns to the effect of using regularized variograms and randomness in the sampling scheme. Consider experiment one (systematic approach) for now, and as an example, Figures B.18, B.19, and B.20. These figures show the confidence intervals of the variogram for point, $\frac{L}{80}$, and $\frac{L}{40}$. The important aspect to see is the consistent dramatic reduction in the spread of the confidence intervals of the variogram. This is the case of experiment one for all the grid resolutions, both in the plain and in the nugget cases. Now consider a similar sequence of variograms for experiment two: Figures D.1, D.2, and D.3. The behavior of the confidence intervals widths is completely different than for experiment one, because the spread does not

get reduced. In fact, the contrary happens because the width actually increases instead of decreasing. The situation is similar for experiment three. Consider Figures D.10, D.11, and D.12. Their behavior is the same as for experiment two (for our comparison purposes). So, we can conclude that the introduction of randomness not only increases the uncertainty of the variogram function estimates, but it spoils the benefits of using a regularized variogram for the estimation of the spatial structure of the field.

Mixed support variograms vs. Two regularized variograms

The issue in question here is for the case when data are available for, say, two support sizes and so the respective regularized variograms are already available as well. From data files at each of the sample support sizes (point, $\frac{L}{80}$, and $\frac{L}{40}$), it is possible to construct a mixed support variogram with two different sample supports and fit the resulting variogram to obtain the point variogram. In particular, experiment number two is more suitable for that comparison because more mixed variogram cases were calculated. In experiment number one, on the other hand, the approach was that of compositing sample sizes. Figures D.13 and D.16 offer a good comparison of this situation, and are an example of several situations of this case, as well as Figures D.14 and D.17 show another example, but with a larger sample support size ($\frac{L}{40}$). Interestingly, the two cases are not so different. In fact, it seems that the use of the mixed support variogram summarizes the data content of the independent regularized variograms used in the mixed suppor variogram. However, one must keep in mind that the cases able of comparison for the mixed vs. two variograms in experiment two are those including the point support observations, either mixed or by itself. There is a very important issue with the use of point variogram observations in addition to any other type of data. I note that the above explanation of the interaction between the point and regularized data is somewhat speculative. Another possible explanation is due to an increase in local minima with the combined data types, thus resulting in greater variation in the regression procedure. This aspect is discussed further below.

Effect of Including all Variable Data in the Fitting

To addres this effect, it is necessary to discuss experiments one and two separately. Recall that for experiment one, the point and $\frac{L}{80}$ data are actually subsamples from the $\frac{L}{40}$ data, thus they typically are less representative of the true point variogram. This is borne out by the sanalysis, se for example the last two columns in Table B.2. This table shows that the confidence intervals for both the sill and range decrease as we move from point to $\frac{L}{80}$, and from $\frac{L}{80}$ to $\frac{L}{40}$. However, as we introduce several data types in the regression, the confidence intervals actually increase. I believe this effect is a consequence of the correspondance of the various data to the true field properties. For instance, the $\frac{L}{40}$ data represent 9% for the $12x12$ case and 16% for the $16x16$ case, while the $\frac{L}{80}$ represent 2.25% for the $12x12$ case and 4% for the $16x16$ case.

A close examination of the data in Table B.2 shows another interesting feature. Not only does combining the various types of data fail to improve the estimation of the point variogram relative to fitting $\frac{L}{40}$ alone, it actually results in a poorer performance from than with fitting the point or $\frac{L}{80}$ data alone. This issue is at first glance rather perplexing. If we weight the parameter estimates for a simulated sampling proportional to the representative field area, we can also estimate a 95% confidence interval. With this approach, assuming unbiased values, we note that it is mathematically impossible that upon combining the data types, the resulting confidence interval is greater than that for the worse case prior to combining. This suggests that the increase in confidence intervals with the addition of the data types of the regression procedure is somehow a consequence of the fitting process. Perhaps related to an increase in the local minima, etc. This issue is outside the scope of the present work, but promises to be a good future research topic.

In experiment two, the situation is somewhat different than for experiment one, since the samples are disjoint sets. In the results for experiment two, recall that there were only 100 simulated samplings, in contrast to 300 simulated samplings of experiment one. As a result, on average, the $\frac{L}{40}$ variogram sampled from a comparatively smaller portion of the field

than the corresponding $\frac{L}{40}$ variogram in experiment one. As an example of how combining various data types influences the fitting process, see Table D.1., especially the rightmost two columns. Here it is observed that the least uncertainty in the estimate of the sill is obtained upon using all the data types simultaneously in the fitting. However, we also note that the estimate of the range is again corrupted by combining the various data types. This is similar to the results for experiment one. We offer again as speculation the two possible factors cited above in this regard, i.e., a disparity in the variability of the two data types, and/or the complexity and subsequent increase in local minima, and possible solutions of the minimization with the contrasting data types.

Gelhar's theorem: *sill* · *integral scale* is a constant

Chapter 6 of [Gelhar, 1993] discusses to considerable extent several key features of geostatistics that are relevant to this work. In particular, on p. 288, the author shows that the product *sill* · *integral scale* should be a constant for regularized variograms at any arbitrary sample support, provided a constraint on the measuring instrument. The property is developed in the book under the following assumptions by the author: stationarity and the results are derived for a one dimensional continuous series. In the book, the author proceeds to say that spatial averaging is a special case of a linear filter or a convolution integral of the form:

$$\gamma_v(h) = \int_{-\infty}^{+\infty} f(x-u)\gamma(h)\, du \tag{6.1}$$

In the case of a moving average (which is how we defined regularization in this dissertation):

$$f(u) = \begin{cases} 1/w, |u| \leq w/2 \\ 0, |u| > w/2 \end{cases} \tag{6.2}$$

Assuming now $\gamma(h)$ to be a stationary process, and using the representation theorem, that:

$$dZ_{\gamma_v} = F(k)dZ_\gamma; \quad F(k) = \int_{-\infty}^{+\infty} e^{-iku} f(u)du \tag{6.3}$$

In the case of a moving average, the transfer function becomes:

$$F(k) = \frac{2\sin(wk/2)}{wk} \tag{6.4}$$

The presentation in the book proceeds now by computing the spectral density function of the filtered (regularized) process using the representation theorem of the form:

$$S_{\gamma_v\gamma_v} = F(k) \cdot F^*(k) \cdot S_{\gamma\gamma} \tag{6.5}$$

Thus, the spectrum of the filtered process is the product of the square of the modulus of the transfer function and the spectrum of the original continuous process. The integral scale presented by Gelhar (1993) for a stationary random process is expressed in terms of the spectral amplitude at zero wave number:

$$integral\ scale\ = \int_0^\infty \rho(s)\ ds = \pi \cdot S(0)/\sigma^2 \tag{6.6}$$

So from the last equation it is clear that the product of the variance and the integral scale remains invariant under the averaging or filtering process. The quantity *sill* · *integral scale* is of great help in determining the actual spatial structure of a data set. For the particular case of this dissertation, a set of tables were prepared to show the behavior of the *sill* · *integral scale* quantity in both experiments 1 and 2, plain and nugget cases. The theoretical value was computed from the original observed field.

In experiment one it is evident that the quantity is better behaved than in experiment one based on the analysis of the mean and standard deviation. In experiment three, the results were much more variable and biased with respect to the theoretical value compared to the experiment one results. From these results, then one can infer that the systematic approach to sampling offers an advantage of preserving the consistency of Gelhar's result. Experiments two and three provide two approaches to randomness in sampling, and so they seem to have a high degree of sparsity for both the nugget and the plain cases.

Histogram Transformations

The most immediate effect of sampling a field with a given sample support is the fact that the statistical distribution of the data will be modified from its original distribution. The main concerns have to do with the reduction of the spread of the data due to the regularization and

Case	12x12	16x16	24x24	32x32	40x40
Point:	481.93	388.31	398.19	395.32	396.42
L/80:	440.44	368.50	412.01	381.79	393.69
L/40:	424.96	376.78	409.08	389.02	-
Mixed:	430.96	371.37	410.02	385.17	-
Case A:	414.15	375.79	408.58	373.24	414.67
Case B:	392.55	382.41	408.91	389.50	-
Case C:	394.62	381.41	408.99	388.46	-
Mean:	425.66	377.80	407.97	386.07	401.59
σ:	30.57	6.80	4.47	7.02	11.41

Table 6.1: Experiment 1, Plain. Table entries are the values of $sill \cdot range$. Theoretical value is 402.47

therefore reduction in variance. There are several methods available in the literature to take care of this problem, of which two were used: the affine correction, and the log transform of the data. In the particular case of this dissertation, both methods were used to construct a de-regularized histogram of the data using a version of the inverse transformation which is easy to derive from the original equations of the method. Since the data fields in this work were very close to normally distributed, then both transformations gave very similar results. A comment to be made about the de-transformed histograms is the fact that the log transform seemed to give less spreaded values. This is of not much surprise because the log transform is designed for data that is more log normally distributed than just normally distributed. In any case, as [Cressie, 1993] says, the research in this particular area is still not giving a good method yet for histogram transformations.

Case	12x12	16x16	24x24	32x32	40x40
Point:	430.74	342.27	346.84	341.23	342.89
L/80:	251.48	317.86	342.17	329.23	335.78
L/40:	242.14	323.63	342.19	334.43	-
Mixed:	245.27	315.98	343.64	331.42	-
Case A:	209.77	317.15	320.83	316.17	344.45
Case B:	167.12	329.22	332.91	331.67	-
Case C:	166.96	326.29	332.20	330.69	-
Mean:	244.78	324.63	337.25	330.65	341.04
σ:	89.39	9.24	9.09	7.53	4.62

Table 6.2: Experiment 1, Nugget. Table entries are the values of $sill \cdot range$. Theoretical value is 353.76

Future Directions

There is still room for future ideas to investigate. For instance, the whole effect of the dispersion variance in the fitting routine was not exhaustively analyzed. A whole dissertation could come out analyzing solely the effect of using dispersion variances for the estimation of the point variogram, as well as modified dispersion variance functions designed to express the relationship in different directions of variability. The whole issue of sampling patterns and programs is a whole science by itself. The book presented in [Thompson and Seber, 1996] is a clear example of this. The area of adaptive sampling is subject of much research and there are many unanswered questions there that could make an interesting topic of research. Finally, at the time the computations were made the reference [Chiles and Delfiner, 1999] had not been published yet, and so many of the ideas presented there were not implemented. In particular, the idea of disjunctive kriging is very appealing to be compared with the two methods presented in this dissertation. A good direction to follow in this line of research could be the use of disjunctive kriging in relation to the sample support effects.

Case	Experiment 2	Experiment 3
Point:	437.06	446.92
L/80:	425.93	469.86
L/40:	460.93	682.76
Mixed: pt & L/80	430.88	432.44
Mixed: pt & L/40	454.01	503.74
Mixed: L/80 & L/40	445.78	628.11
pt + L/80:	424.58	425.87
pt + L/40:	489.79	620.58
L/80 + L/40:	469.04	552.94
All variograms:	447.37	526.25
Mean:	448.54	528.94
σ:	20.67	90.23

Table 6.3: Experiments 2 and 3, Plain. Table entries are the values of $sill \cdot range$. Theoretical value is 402.47

Case	Experiment 2	Experiment 3
Point:	353.04	418.82
L/80:	381.44	403.69
L/40:	371.37	541.57
Mixed: pt & L/80	368.46	371.44
Mixed: pt & L/40	369.57	462.30
Mixed: L/80 & L/40	369.61	526.21
pt + L/80:	351.84	332.03
pt + L/40:	375.01	497.29
L/80 + L/40:	370.94	444.21
All variograms:	354.87	423.69
Mean:	366.61	442.13
σ:	9.96	66.58

Table 6.4: Experiments 2 and 3, Nugget. Table entries are the values of $sill \cdot range$. Theoretical value is 353.76

Numerical Expressions for $\bar{\gamma}(h, v, V)$

The equation of main interest to the purposes of this paper is the following:

$$\bar{\gamma}[v(x), V(x+h)] = \frac{1}{vV} \int_{v(x)} \int_{V(x+h)} \gamma(\vec{r} - \vec{s})\, d\vec{r}\, d\vec{s} \tag{A.1}$$

Equation (A.1) applies to one, two, or three dimensions. The integral symbol represents either a single, double, or triple integration, depending on the number of dimensions of the problem. This equation represents the general formulation of the problem. The following sections explain in detail the transformation of this integral into an expression that can be easily programmed in a computer using a numerical integration method.

A.1 1-Dimensional case

Equation (A.1) is written as the following one dimensional expression:

$$\bar{\gamma}[l(x), L(x+h)] = \frac{1}{lL} \int_{l} \int_{L} \gamma(r - s)\, dr\, ds \tag{A.2}$$

From this point on, the method explained in [Journel and Huijbregts, 1978] is followed: use two index functions $k(y)$ and $k'(y')$ such that:

$$k(y) = \begin{cases} 1 & : \quad \vec{y} \in l \\ 0 & : \quad otherwise \end{cases} \qquad k'(y') = \begin{cases} 1 & : \quad \vec{y'} \in L \\ 0 & : \quad otherwise \end{cases}$$

The function definitions above imply that the following conditions hold:

$$\int_{-\infty}^{+\infty} k(y)\, dy = l \qquad \int_{-\infty}^{+\infty} k'(y')\, dy' = L$$

With these results in mind, equation (A.2) can be rewritten as:

$$\bar{\gamma}[l(x), L(x+h)] = \frac{1}{lL} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} k(r) \cdot k'(s) \cdot \gamma(r-s)\, dr\, ds \tag{A.3}$$

Define now the function K_{mix} , the convolution of the indicator functions k and k', like this (doing a change of variable of the form $u + h = r - s$):

$$K_{mix}(u) = k * \check{k}' = \int_{-\infty}^{+\infty} k(u+y') \cdot k'(y')\, dy' \tag{A.4}$$

Equation (A.4) can be now used to simplify equation (A.3) as follows:

$$\bar{\gamma}[l(x), L(x+h)] = \frac{1}{l \cdot L} \int_{-\infty}^{+\infty} K_{mix}(u)\gamma(u+h)\, du \tag{A.5}$$

The double integral was then reduced to a single integral by using the indicator function. This procedure is called the *Cauchy-Gauss method* and is explained for a more simple case where the two support volumes are the same in [Journel and Huijbregts, 1978]. Such method is useful only when the specific form of the function $K_{mix}(u)$ is known a priori. Otherwise, it is of little use. In order to derive a specific form of the function $K_{mix}(u)$, it is important to realize that it represents the fact that we have two indicator functions for two different support volumes. To see this more clearly, Figure A.1 shows the mixed support arrangement

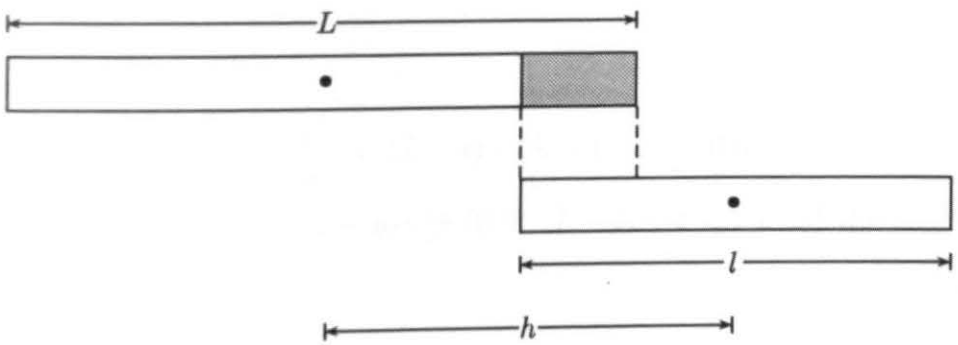


Figure A.1: 1-Dimensional mixed support overlaping length

that makes this more clear. A function that combines the two support volumes (lengths) is then a function that gives the value of the overlapping length when one end of the separation vector points to support l and the other end of the separation vector points to support L.

With this in mind, the definition of such a function is straightforward. Taking as a reference point the center of each support volume, the *mixed indicator* function becomes:

$$K_{mix}(u) = \begin{cases} \frac{1}{2}(L+l) - |u| & : \quad \frac{1}{2}(L-l) \leq |u| \leq \frac{1}{2}(L+l) \\ l & : \quad 0 \leq |u| \leq \frac{1}{2}(L-l) \\ 0 & : \quad |u| \geq \frac{1}{2}(L+l) \end{cases} \tag{A.6}$$

Using equation (A.6), the integral definition in equation (A.5) becomes the sum of three integrals, as follows:

$$\bar{\gamma}[l(x), L(x+h)] = \tag{A.7}$$

$$\begin{aligned} &= \frac{1}{l \cdot L} \int_{-\infty}^{+\infty} K_{mix}(u)\gamma(u+h)\, du \\ &= \frac{1}{l \cdot L} \left(\int_{-0.5(L+l)}^{-0.5(L-l)} [0.5 \cdot (L+l) + u]\, \gamma(u+h)\, du \right. \\ &\quad + \int_{-0.5(L-l)}^{+0.5(L-l)} l \cdot \gamma(u+h)\, du \\ &\quad \left. + \int_{+0.5(L-l)}^{+0.5(L+l)} [0.5 \cdot (L+l) - u]\, \gamma(u+h)\, du \right) \end{aligned} \tag{A.8}$$

Equation (A.7) can be further simplified by doing appropriate changes of variable. The details are rather tedious and are not included in this presentation. The final results of such simplification are shown below (equation A.9):

$$\bar{\gamma}[l(x), L(x+h)] = \tag{A.9}$$

$$\begin{aligned} &= \frac{1}{l \cdot L} \int_0^1 \left\{ l \cdot (L-l)\gamma\left[(L-l) \cdot (y-0.5) + h\right] \right. \\ &\quad \left. + l^2 \cdot (1-y)\gamma\left[\pm\,(0.5 \cdot (L-l) + l \cdot y) + h\right] \right\} dy \end{aligned} \tag{A.10}$$

Where $\pm$ means :

$$\gamma\left(\pm a \pm b\right) = \gamma(a+b) + \gamma(-a+b) + \gamma(a-b) + \gamma(-a-b)$$

The last equation, (A.9), is in a very convenient way for programming purposes. The fact that the integration interval is between 0 and 1 makes it very appropriate for a solution method such as Gaussian Quadrature [Abramowitz and Stegun, 1970, Press et al., 1992].

A.2 2-Dimensional case

The two dimensional case can be derived based on the one dimensional case, if we consider that the support volumes (areas in this case) are assumed to be rectangular, with the sides parallel to the coordinate axis. Figure A.2 shows a possible configuration of two supports

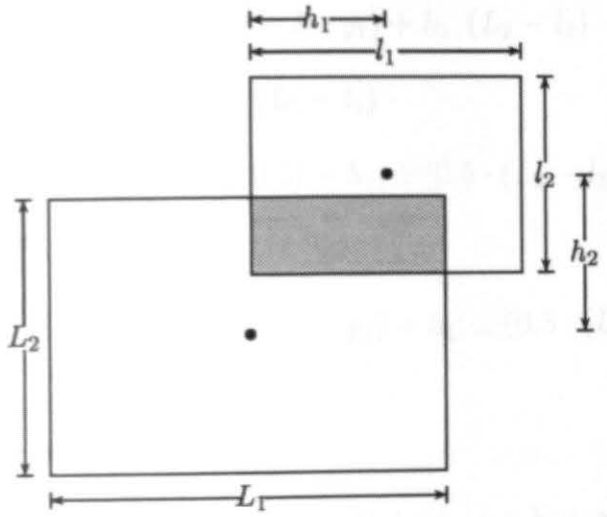


Figure A.2: 2-Dimensional mixed support overlaping area

defined by two rectangles, one with area l_1 by l_2, and the other one with area L_1 by L_2. The K_{mix} function represents the amount of overlap between the two support volumes (areas in this case), which is represented in the figure by the squared region. The function K_{mix} can be written as follows (taking as a reference the center of the rectangles):

$$K_{mix}(u_1, u_2) = m_1 \cdot m_2 \tag{A.11}$$

$$\text{Where } m_i = \begin{cases} \frac{1}{2}(L_i + l_i) - |u_i| & : \quad \frac{1}{2}(L_i - l_i) \leq |u_i| \leq \frac{1}{2}(L_i + l_i) \\ l_i & : \quad 0 \leq |u_i| \leq \frac{1}{2}(L_i - l_i) \\ 0 & : \quad |u_i| \geq \frac{1}{2}(L_i + l_i) \end{cases}$$

If we let $a = l_1 \cdot l_2$, and $A = L_1 \cdot L_2$, and using the results obtained for the one dimensional case, the equivalent of equation (A.9), in the two dimensional case, becomes:

$$\bar{\gamma}[a(\vec{x}), A(\vec{x} + \vec{h})] = \tag{A.12}$$

$$
\begin{aligned}
&= \frac{1}{a \cdot A} \int_{-0.5(L_1+l_1)}^{+0.5(L_1+l_1)} \int_{-0.5(L_2+l_2)}^{+0.5(L_2+l_2)} K_{mix}(u_1, u_2)\gamma(u_1 + h_1, u_2 + h_2)\, du_1\, du_2 \\
&= \frac{1}{a \cdot A} \int_0^1 \int_0^1 \Big\{ l_2 \cdot (L_2 - l_2) \cdot \Big[l_1 \cdot (L_1 - l_1) \cdot \\
&\quad \gamma\left[(L_1 - l_1) \cdot (y_1 - 0.5) + h_1, (L_2 - l_2) \cdot (y_2 - 0.5) + h_2\right] \\
&\quad + l_1^2 \cdot (1 - y_1) \cdot \\
&\quad \gamma\left[\pm\left(0.5 \cdot (L_1 - l_1) + l_1 \cdot y_1\right) + h_1, (L_2 - l_2) \cdot (y_2 - 0.5) + h_2\right] \Big] \\
&\quad + l_2^2 \cdot (1 - y_2) \cdot \Big[l_1 \cdot (L_1 - l_1) \cdot \\
&\quad \gamma\left[(L_1 - l_1) \cdot (y_1 - 0.5) + h_1, \pm\left(0.5 \cdot (L_2 - l_2) + l_2 \cdot y_2\right) + h_2\right] \\
&\quad + l_1^2 \cdot (1 - y_1) \cdot \\
&\quad \gamma\left[\pm\left(0.5 \cdot (L_1 - l_1) + l_1 \cdot y_1\right) + h_1, \pm\left(0.5 \cdot (L_2 - l_2) + l_2 \cdot y_2\right) + h_2\right] \Big] \Big\}\, dy_1\, dy_2
\end{aligned}
\tag{A.13}
$$

Where $\pm$ means :

$$
\gamma\left(\pm a \pm b\right) \;=\; \gamma(a+b) + \gamma(-a+b) + \gamma(a-b) + \gamma(-a-b)
$$

A.3 3-Dimensional case

The three dimensional case is not necessary for the analysis in the paper. However, it is included here for completeness. The equations are straightforward combining the one- and two-dimensional cases, considering that the sample support volumes are rectangular blocks, whose volumes are defined by: $v = l_1 \cdot l_2 \cdot l_3$, and $V = L_1 \cdot L_2 \cdot L_3$. The K_{mix} function is an extension of the two-dimensional case, and is defined by: $K_{mix}(u_1, u_2, u_3) = m_1 \cdot m_2 \cdot m_3$, with m_i defined by equation (A.11). The amount of work required to reduce the equations to integrals in the interval $[0, 1]$ is extensive, and so only the final equations are shown below:

$$
\begin{aligned}
\bar{\gamma}[v(\vec{x}), V(\vec{x} + \vec{h})] \;&= \\
&= \frac{1}{v \cdot V} \int_{-0.5(L_1+l_1)}^{+0.5(L_1+l_1)} \int_{-0.5(L_2+l_2)}^{+0.5(L_2+l_2)} \int_{-0.5(L_3+l_3)}^{+0.5(L_3+l_3)} K_{mix}(u_1, u_2, u_3) \cdot \\
&\quad \gamma(u_1 + h_1, u_2 + h_2, u_3 + h_3)\, du_1\, du_2\, du_3
\end{aligned}
\tag{A.14}
$$

$$= \frac{1}{v \cdot V} \int_0^1 \int_0^1 \int_0^1 \Big\{ l_2 \cdot (L_2 - l_2) \cdot l_3 \cdot (L_3 - l_3) \cdot \Big[l_1 \cdot (L_1 - l_1) \cdot$$
$$\gamma\,[(L_1 - l_1) \cdot (y_1 - 0.5) + h_1,$$
$$(L_2 - l_2) \cdot (y_2 - 0.5) + h_2, (L_3 - l_3) \cdot (y_3 - 0.5) + h_3]$$
$$+ l_1^2 \cdot (1 - y_1) \cdot$$
$$\gamma\,[\pm\,(0.5 \cdot (L_1 - l_1) + l_1 \cdot y_1) + h_1,$$
$$(L_2 - l_2) \cdot (y_2 - 0.5) + h_2, (L_3 - l_3) \cdot (y_3 - 0.5) + h_3] \Big]$$
$$+ l_3 \cdot (L_3 - l_3) \cdot l_2^2 \cdot (1 - y_2) \cdot \Big[l_1 \cdot (L_1 - l_1) \cdot$$
$$\gamma\,[(L_1 - l_1) \cdot (y_1 - 0.5) + h_1,$$
$$\pm\,(0.5 \cdot (L_2 - l_2) + l_2 \cdot y_2) + h_2, (L_3 - l_3) \cdot (y_3 - 0.5) + h_3]$$
$$+ l_1^2 \cdot (1 - y_1) \cdot$$
$$\gamma\,[\pm\,(0.5 \cdot (L_1 - l_1) + l_1 \cdot y_1) + h_1,$$
$$(L_2 - l_2) \cdot (y_2 - 0.5) + h_2, \pm\,(0.5 \cdot (L_3 - l_3) + l_3 \cdot y_3) + h_3] \Big]$$
$$+ l_3^2 \cdot (1 - y_3) \cdot l_2 \cdot (L_2 - l_2) \cdot \Big[l_1 \cdot (L_1 - l_1) \cdot$$
$$\gamma\,[(L_1 - l_1) \cdot (y_1 - 0.5) + h_1,$$
$$(L_2 - l_2) \cdot (y_2 - 0.5) + h_2, \pm\,(0.5 \cdot (L_3 - l_3) + l_3 \cdot y_3) + h_3]$$
$$+ l_1^2 \cdot (1 - y_1) \cdot$$
$$\gamma\,[\pm\,(0.5 \cdot (L_1 - l_1) + l_1 \cdot y_1) + h_1,$$
$$(L_2 - l_2) \cdot (y_2 - 0.5) + h_2, \pm\,(0.5 \cdot (L_3 - l_3) + l_3 \cdot y_3) + h_3] \Big]$$
$$+ l_2^2 \cdot (1 - y_2) \cdot l_3^2 \cdot (1 - y_3) \cdot \Big[l_1 \cdot (L_1 - l_1) \cdot$$
$$\gamma\,[(L_1 - l_1) \cdot (y_1 - 0.5) + h_1,$$
$$\pm\,(0.5 \cdot (L_2 - l_2) + l_2 \cdot y_2) + h_2, \pm\,(0.5 \cdot (L_3 - l_3) + l_3 \cdot y_3) + h_3]$$
$$+ l_1^2 \cdot (1 - y_1) \cdot$$
$$\gamma\,[\pm\,(0.5 \cdot (L_1 - l_1) + l_1 \cdot y_1) + h_1,$$
$$\pm\,(0.5 \cdot (L_2 - l_2) + l_2 \cdot y_2) + h_2, \pm\,(0.5 \cdot (L_3 - l_3) + l_3 \cdot y_3) + h_3] \Big]$$

$$\Big\} dy_1\, dy_2\, dy_3$$

Where $\pm$ means :

$$\gamma(\pm a \pm b) = \gamma(a+b) + \gamma(-a+b) + \gamma(a-b) + \gamma(-a-b) \tag{A.15}$$

Numerical Experiment 1 - Plain

Parameter	Mean	Bias	%Bias	S.D.	%S.D.	C.V.	Low C.I.	High C.I.	S.D. C.I.	Low C.I.	High C.I.
Point:											
Sill	7.277	0.054	0.747	1.050	14.432	0.144	7.256	7.298	0.030	5.479	9.310
Range	66.226	10.505	15.863	17.307	26.133	0.261	-238.056	370.509	354.231	9.740	103.192
L/80:											
Sill	7.370	0.147	1.995	1.053	14.291	0.143	7.241	7.499	0.999	5.717	9.564
Range	59.761	4.040	6.760	32.252	53.969	0.540	-84.691	204.211	253.831	6.094	103.871
L/40:											
Sill	7.136	0.087	1.226	1.055	14.780	0.148	6.979	7.292	0.377	5.156	9.305
Range	59.552	3.831	6.433	31.221	52.427	0.524	-38.529	157.633	154.539	6.432	99.349
Mixed:											
Sill	7.267	0.044	0.611	1.039	14.294	0.143	7.149	7.385	0.365	5.470	9.385
Range	59.304	3.583	6.041	31.511	53.136	0.531	-54.581	173.188	181.693	6.269	100.990
Case A:											
Sill	7.336	0.113	1.536	0.992	13.519	0.135	7.323	7.349	0.020	5.814	9.410
Range	56.455	0.734	1.301	34.186	60.554	0.606	-6.522	119.433	157.484	6.477	107.469
Case B:											
Sill	7.271	0.048	0.664	0.966	13.291	0.133	7.261	7.281	0.018	5.801	9.374
Range	53.988	1.733	3.210	32.425	60.060	0.601	22.591	85.385	176.049	6.425	103.796
Case C:											
Sill	7.286	0.063	0.865	0.961	13.196	0.132	7.278	7.294	0.013	5.840	9.383
Range	54.161	1.560	2.880	32.524	60.051	0.601	33.295	75.027	104.929	6.420	104.453

Table B.1: Experiment 1, Plain 12 x 12, sill = 7.223, range = 3*55.721. Confidence Intervals (C.I.) are at the 95% level

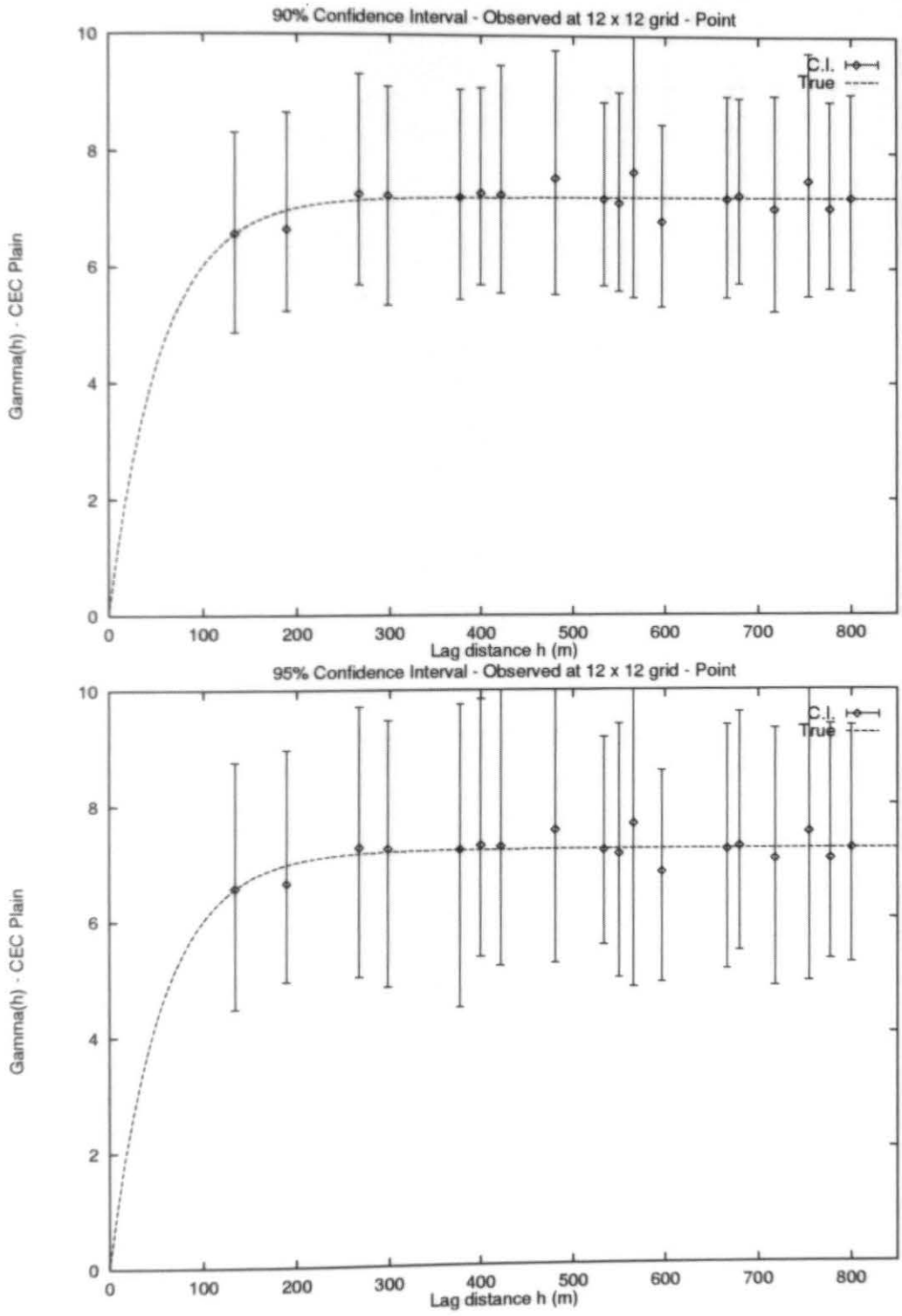


Figure B.1: Numerical Experiment 1 Plain, 12x12 - Obs. Point

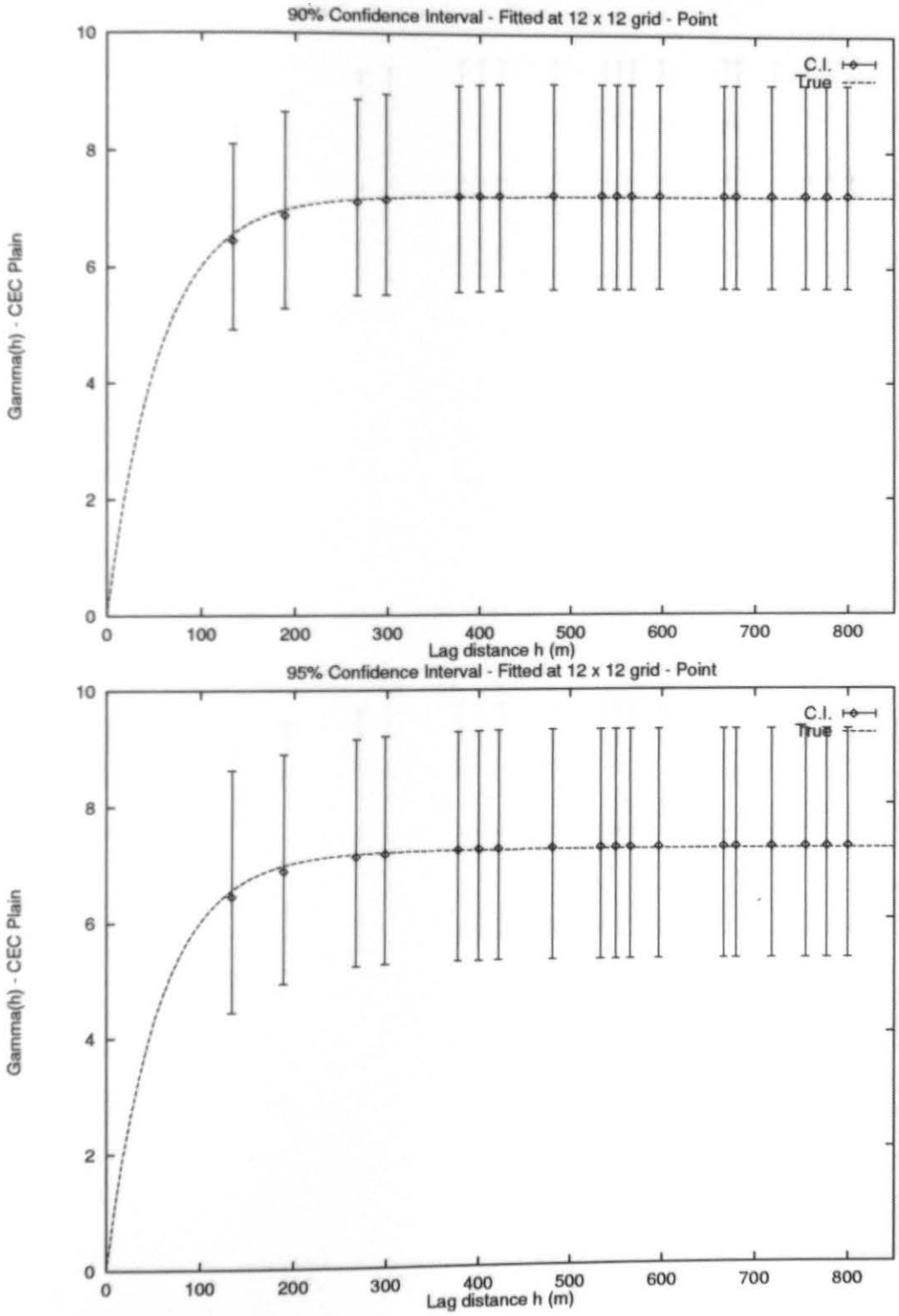


Figure B.2: Numerical Experiment 1 Plain, 12x12 - Fit. Point

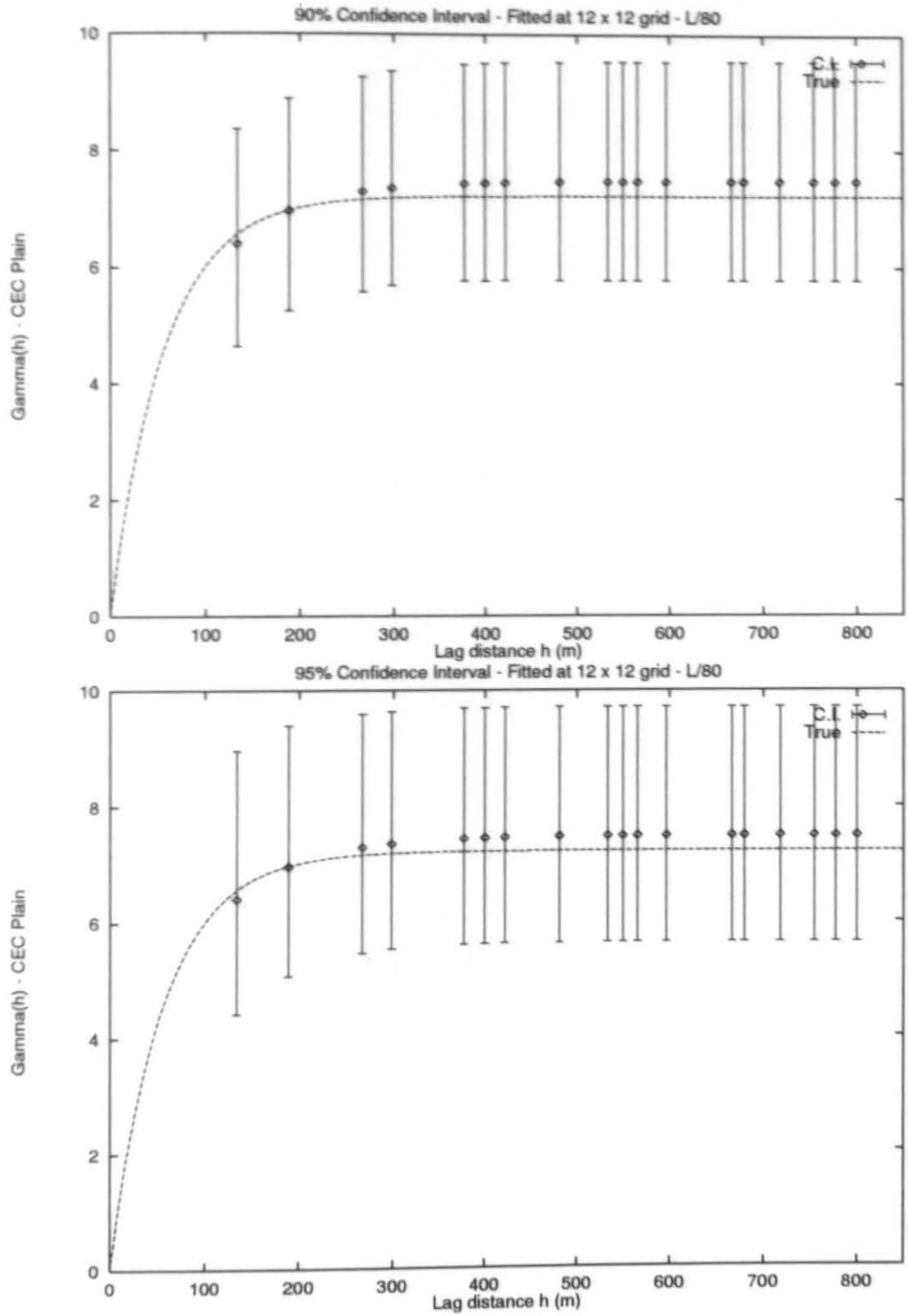


Figure B.3: Numerical Experiment 1 Plain, 12x12 - L/80

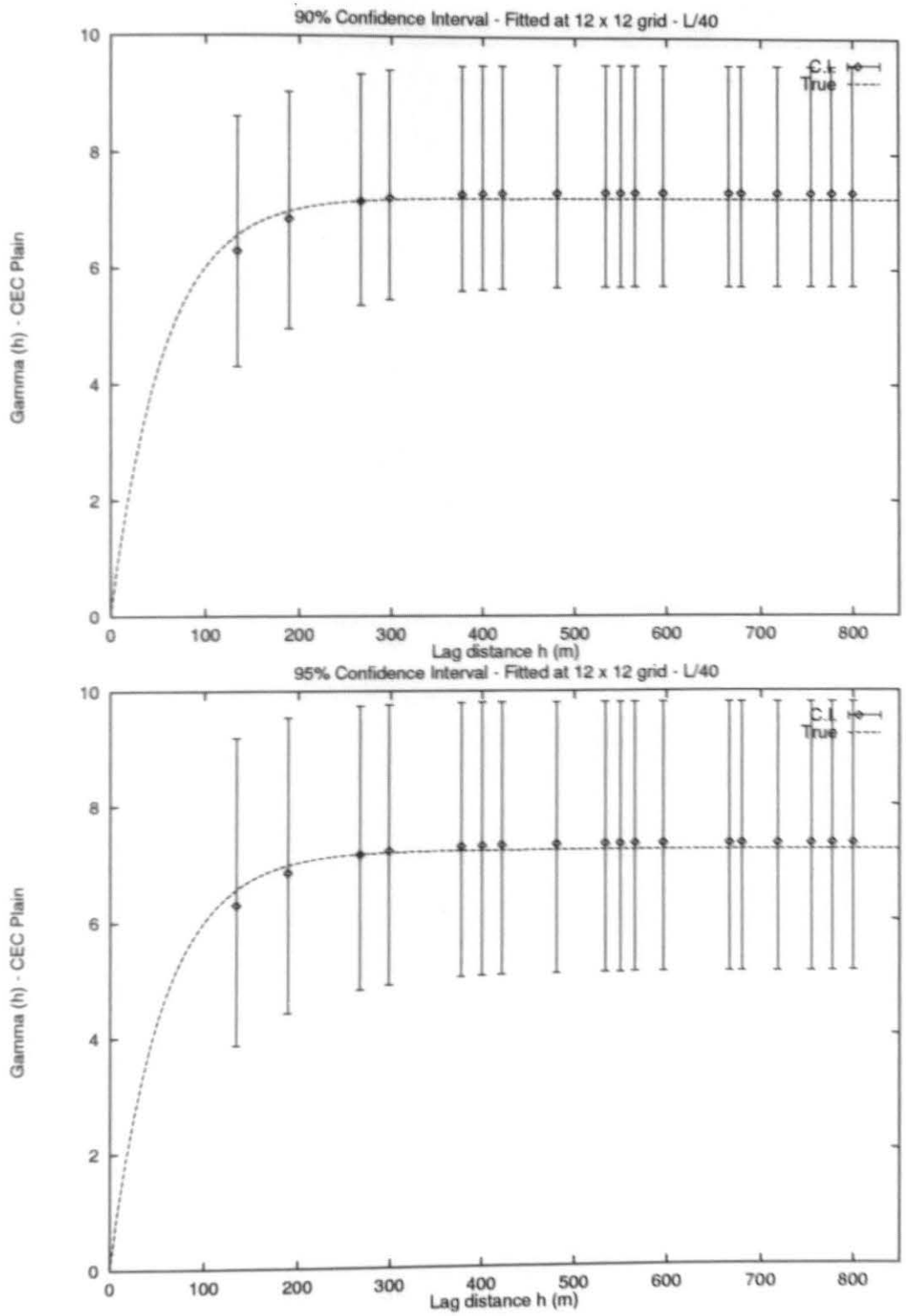


Figure B.4: Numerical Experiment 1 Plain, 12x12 - L/40

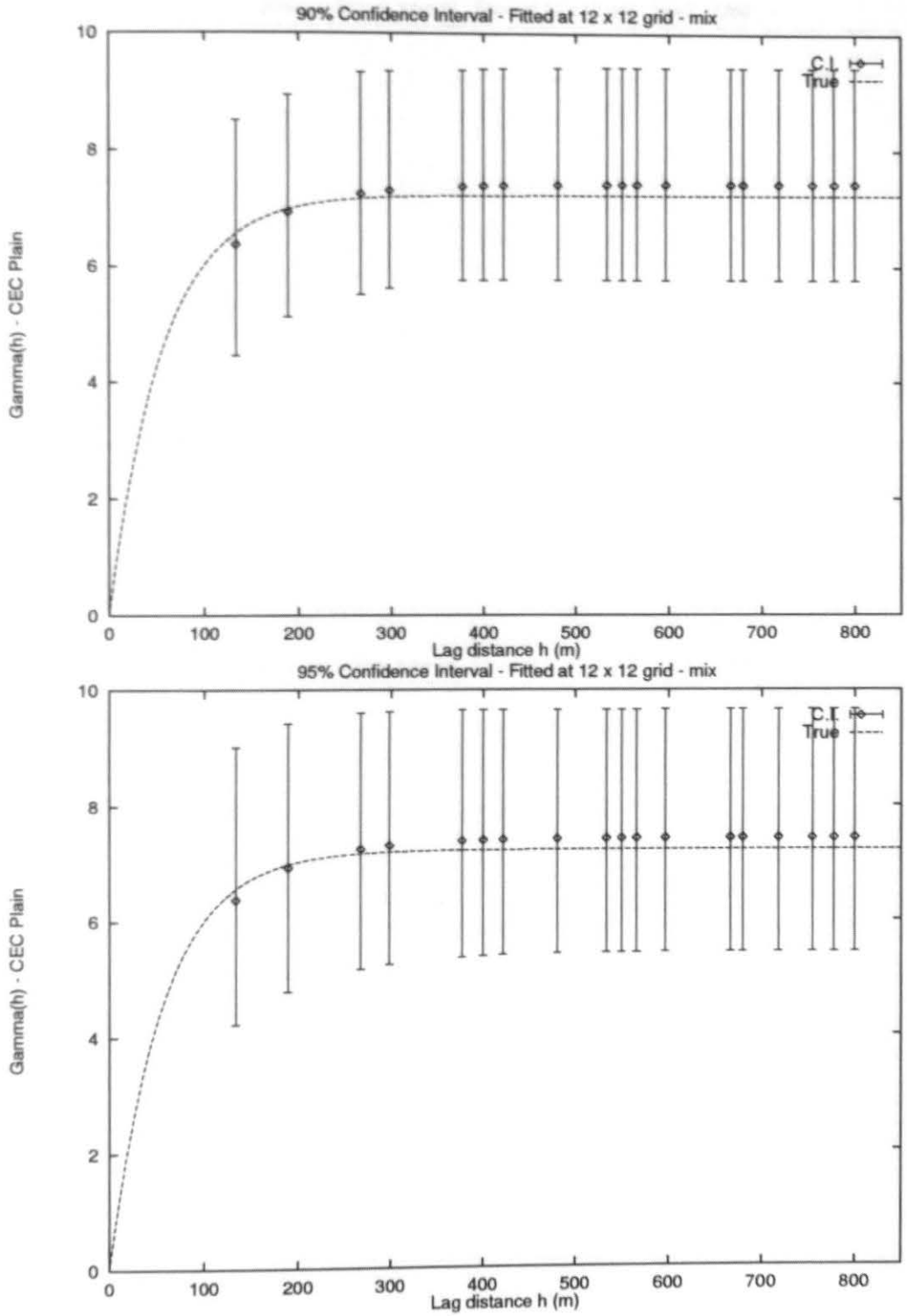


Figure B.5: Numerical Experiment 1 Plain, 12x12 - mix

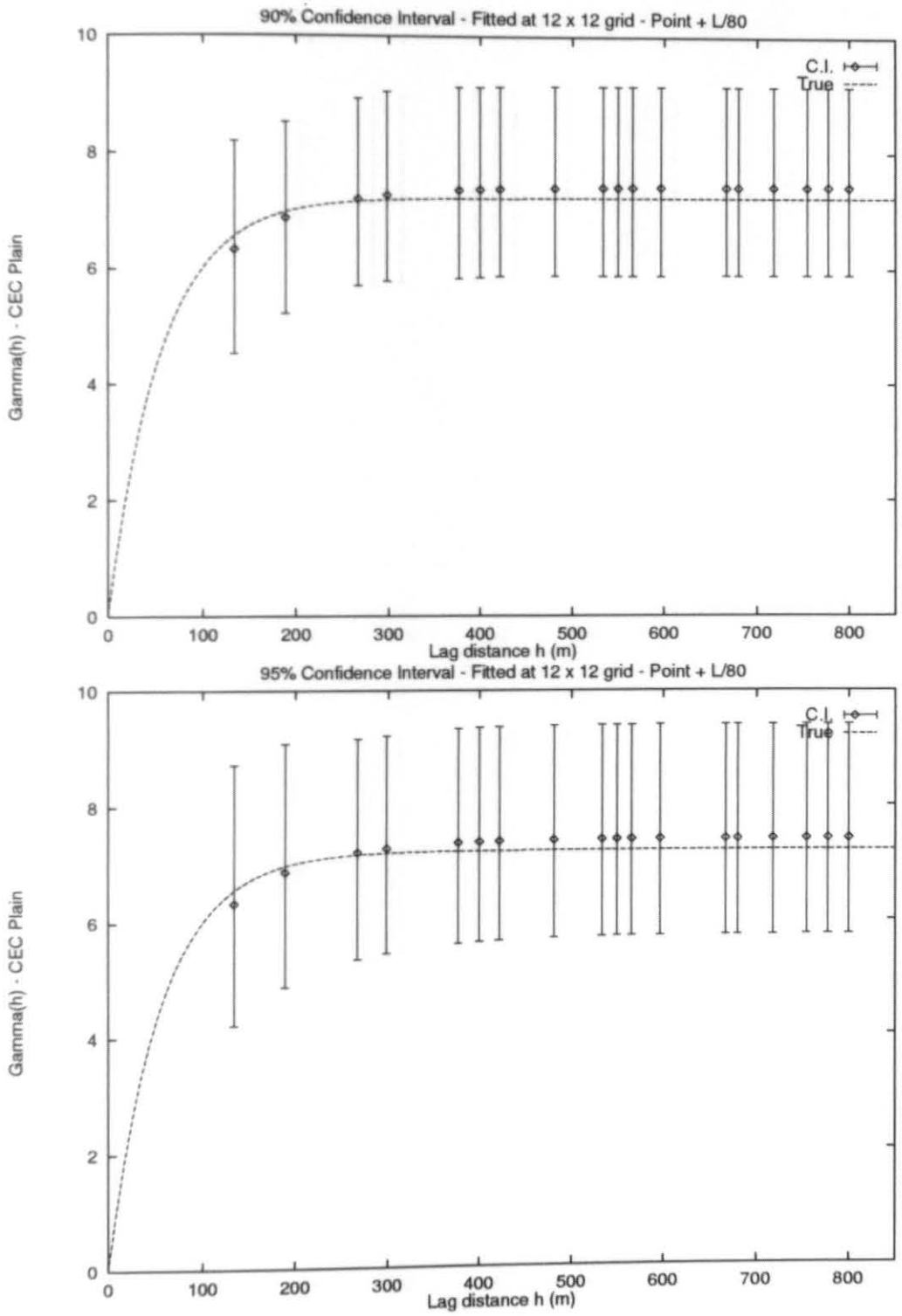


Figure B.6: Numerical Experiment 1 Plain, 12x12 - point + L/80

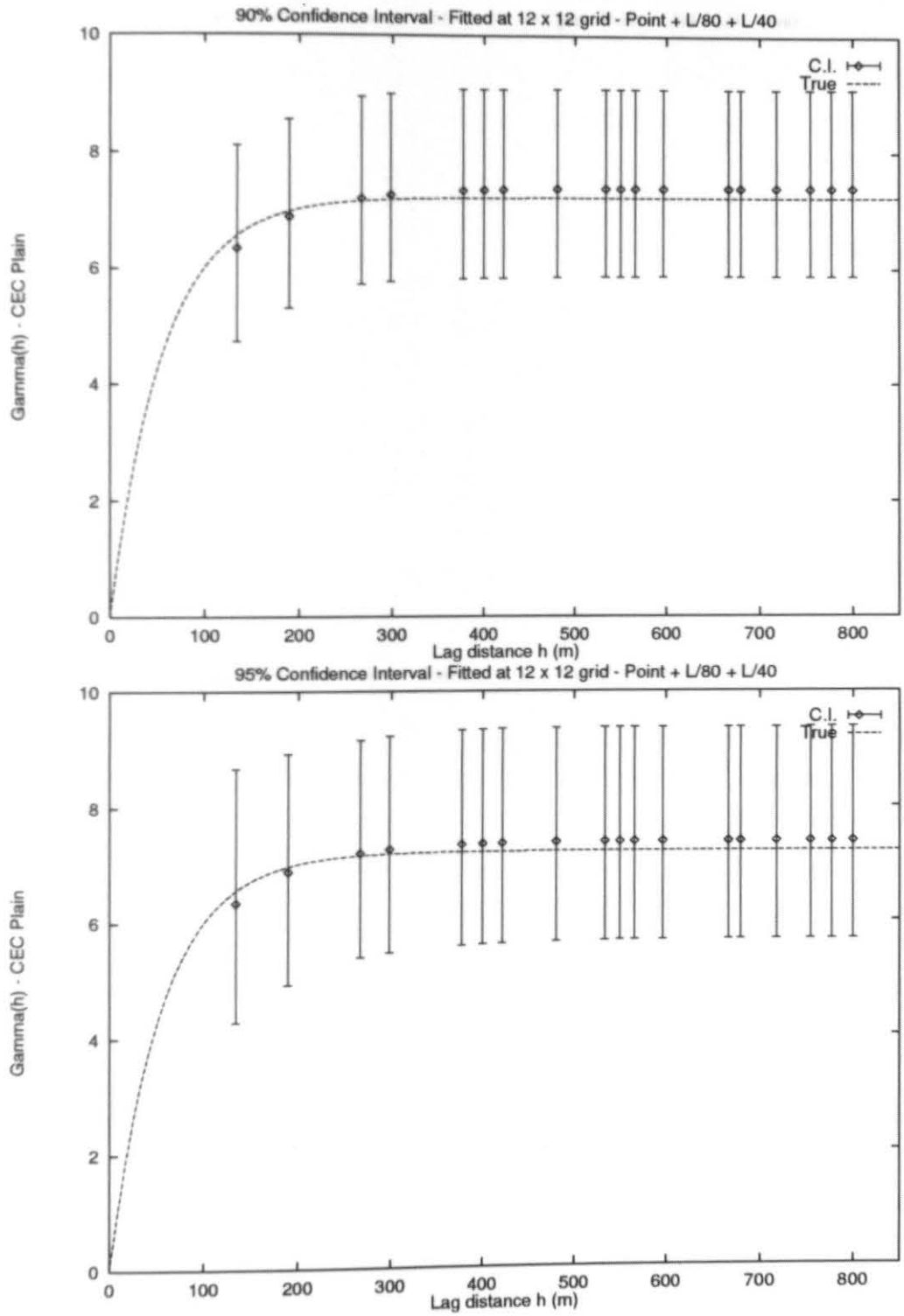


Figure B.7: Numerical Experiment 1 Plain, 12x12 - point + L/80 + L/40

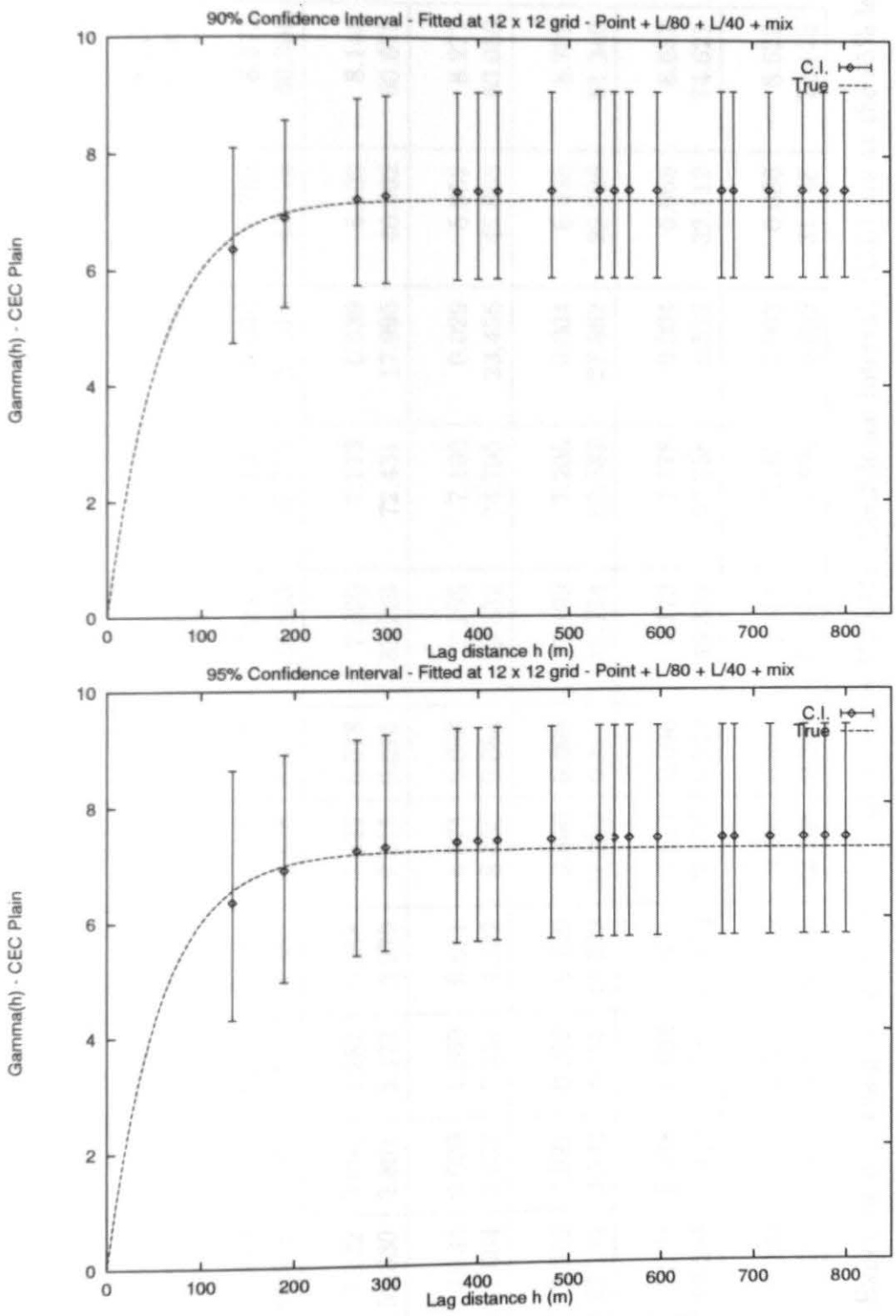


Figure B.8: Numerical Experiment 1 Plain, 12x12 - point + L/80 + L/40 + mix

Parameter	Mean	Bias	%Bias	S.D.	%S.D.	C.V.	Low C.I.	High C.I.	S.D. C.I.	Low C.I.	High C.I.
Point:											
Sill	7.157	0.066	0.918	0.784	10.956	0.110	7.153	7.162	0.006	5.780	8.825
Range	54.256	1.465	2.701	8.855	16.320	0.163	4.136	104.375	56.203	37.736	71.867
L/80:											
Sill	7.110	0.113	1.583	0.688	9.672	0.097	7.093	7.128	0.020	5.762	8.372
Range	51.828	3.893	7.511	4.860	9.378	0.094	23.883	79.773	31.315	43.773	60.300
L/40:											
Sill	7.132	0.091	1.282	0.558	7.827	0.078	7.090	7.173	0.039	6.102	8.145
Range	52.830	2.891	5.473	3.970	7.515	0.075	33.229	72.431	17.995	46.362	60.678
Mixed:											
Sill	7.133	0.090	1.269	0.611	8.571	0.086	7.105	7.160	0.029	5.952	8.277
Range	52.064	3.657	7.024	4.392	8.436	0.084	29.332	74.795	23.456	45.065	60.050
Case A:											
Sill	7.202	0.021	0.292	0.709	9.850	0.098	7.199	7.205	0.004	6.088	8.702
Range	52.178	3.543	6.791	16.322	31.282	0.313	40.774	63.582	27.862	22.726	81.240
Case B:											
Sill	7.175	0.048	0.668	0.691	9.631	0.096	7.173	7.178	0.004	6.065	8.637
Range	53.298	2.423	4.546	12.152	22.799	0.228	49.338	57.258	8.522	32.119	74.621
Case C:											
Sill	7.183	0.040	0.553	0.682	9.492	0.095	7.181	7.185	0.003	6.086	8.622
Range	53.099	2.622	4.937	12.090	22.768	0.228	50.304	55.894	5.999	31.748	74.125

Table B.2: Experiment 1, Plain, 16 x 16, sill = 7.223, range = 3*55.721. Confidence Intervals (C.I.) are at the 95% level

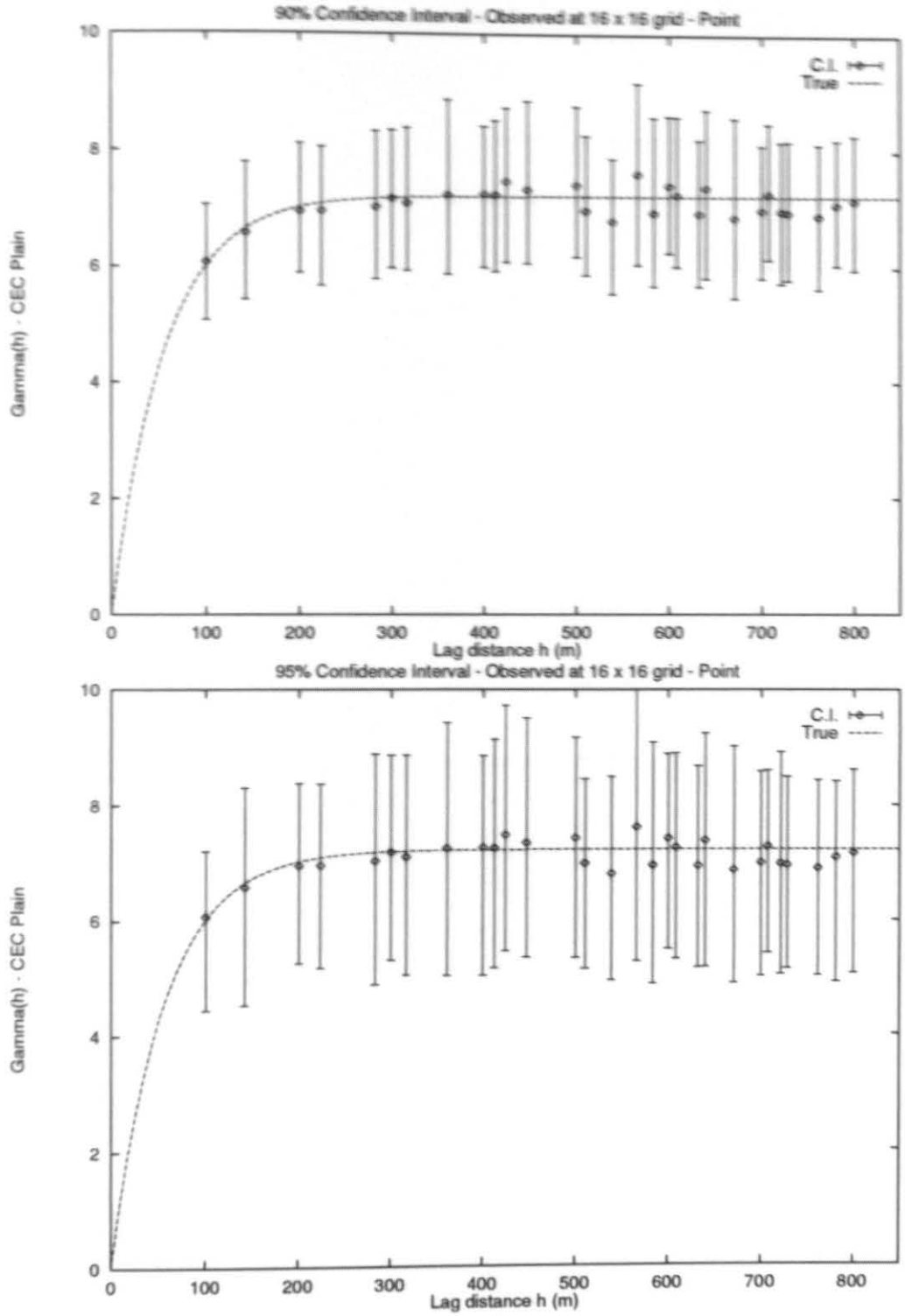


Figure B.9: Numerical Experiment 1 Plain, 16x16 - Obs. Point

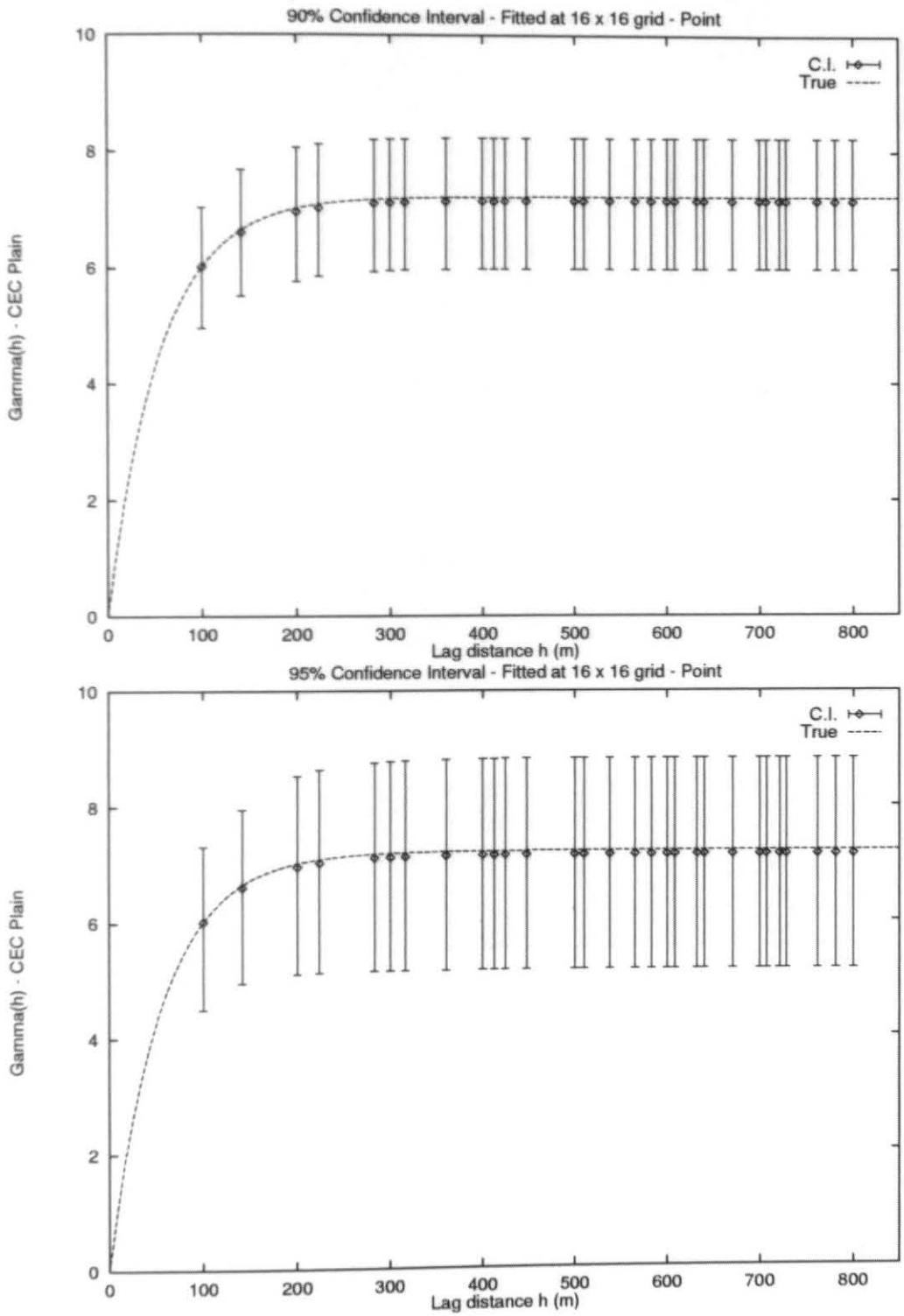


Figure B.10: Numerical Experiment 1 Plain, 16x16 - Fit. Point

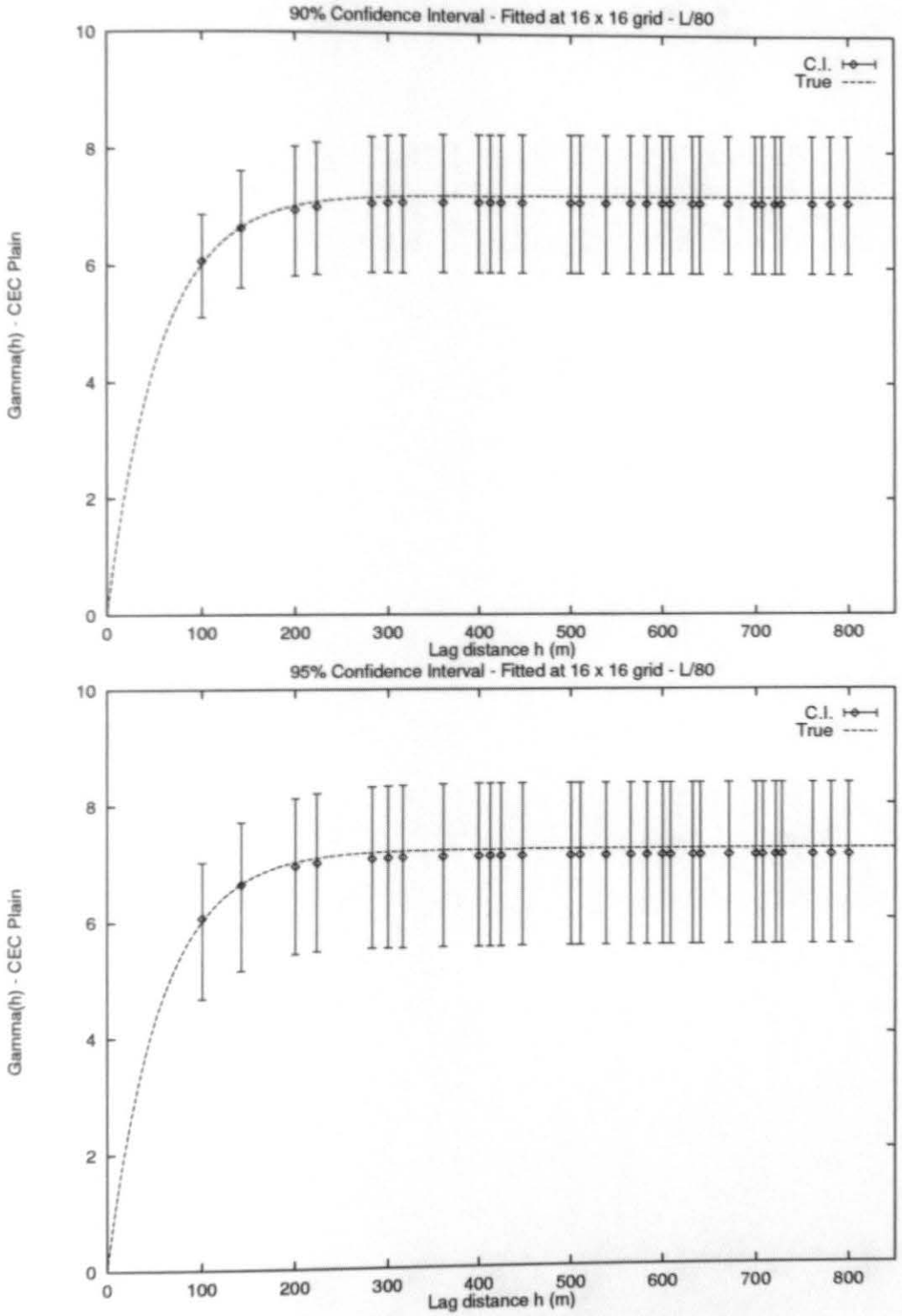


Figure B.11: Numerical Experiment 1 Plain, 16x16 - L/80

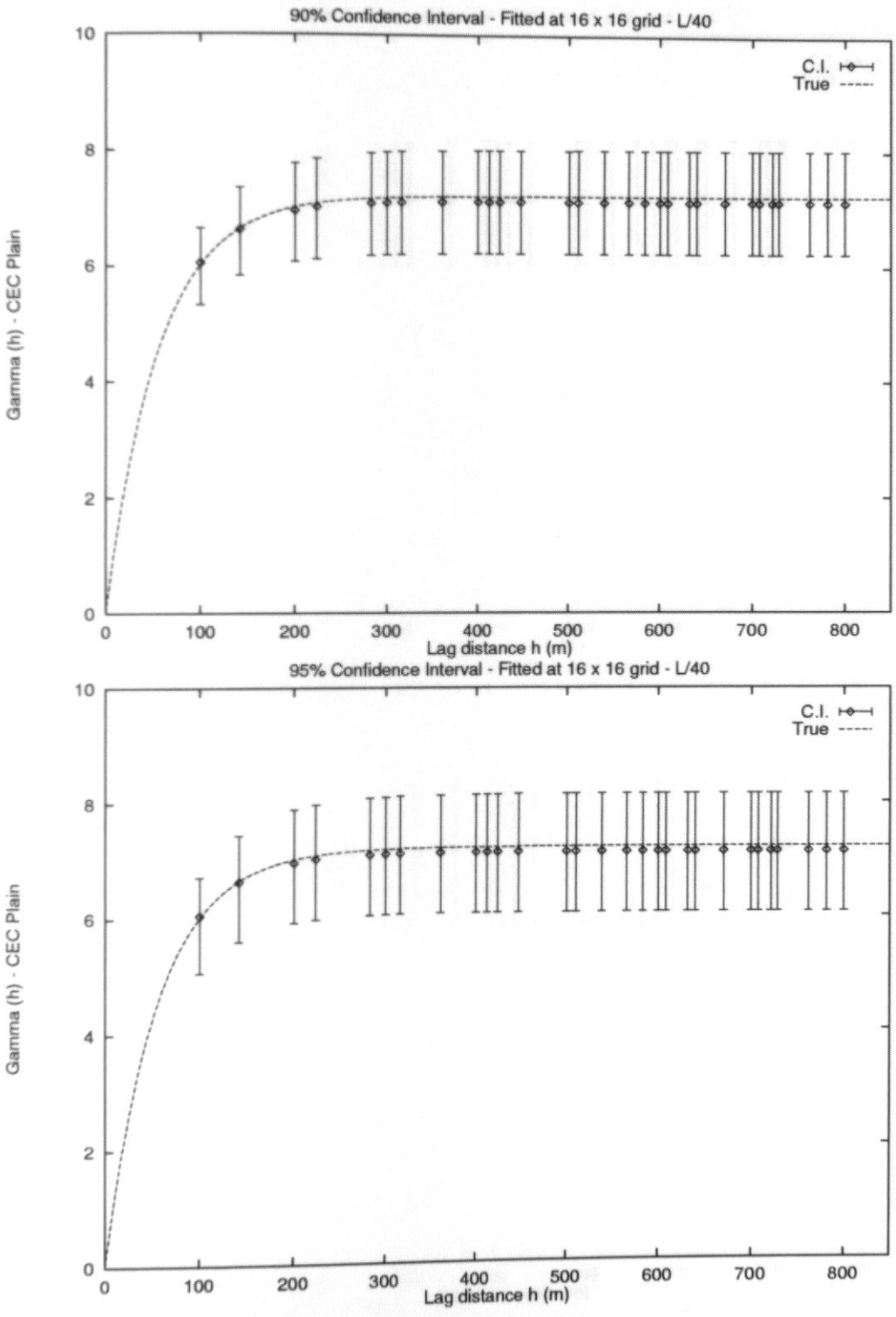


Figure B.12: Numerical Experiment 1 Plain, 16x16 - L/40

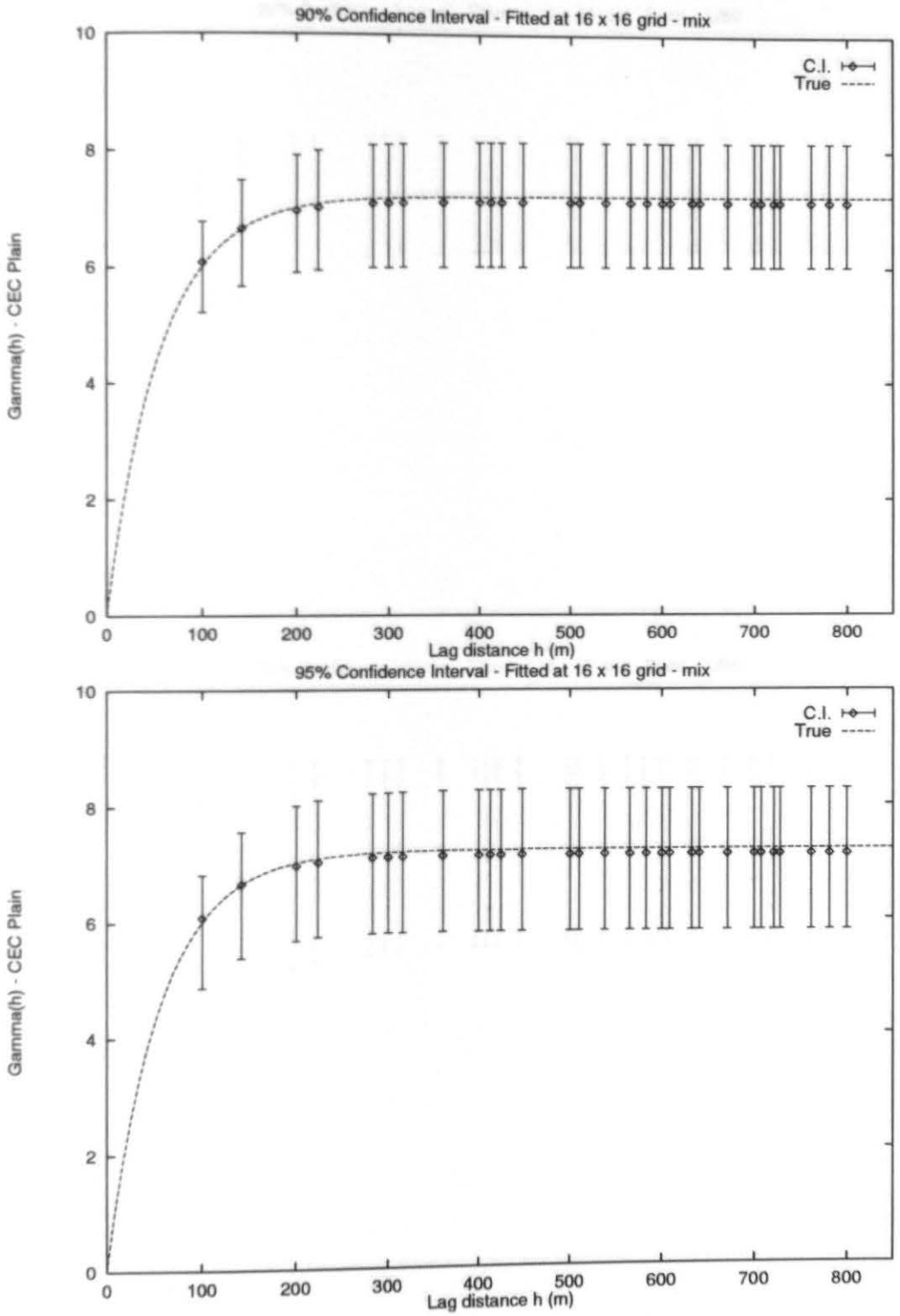


Figure B.13: Numerical Experiment 1 Plain, 16x16 - mix

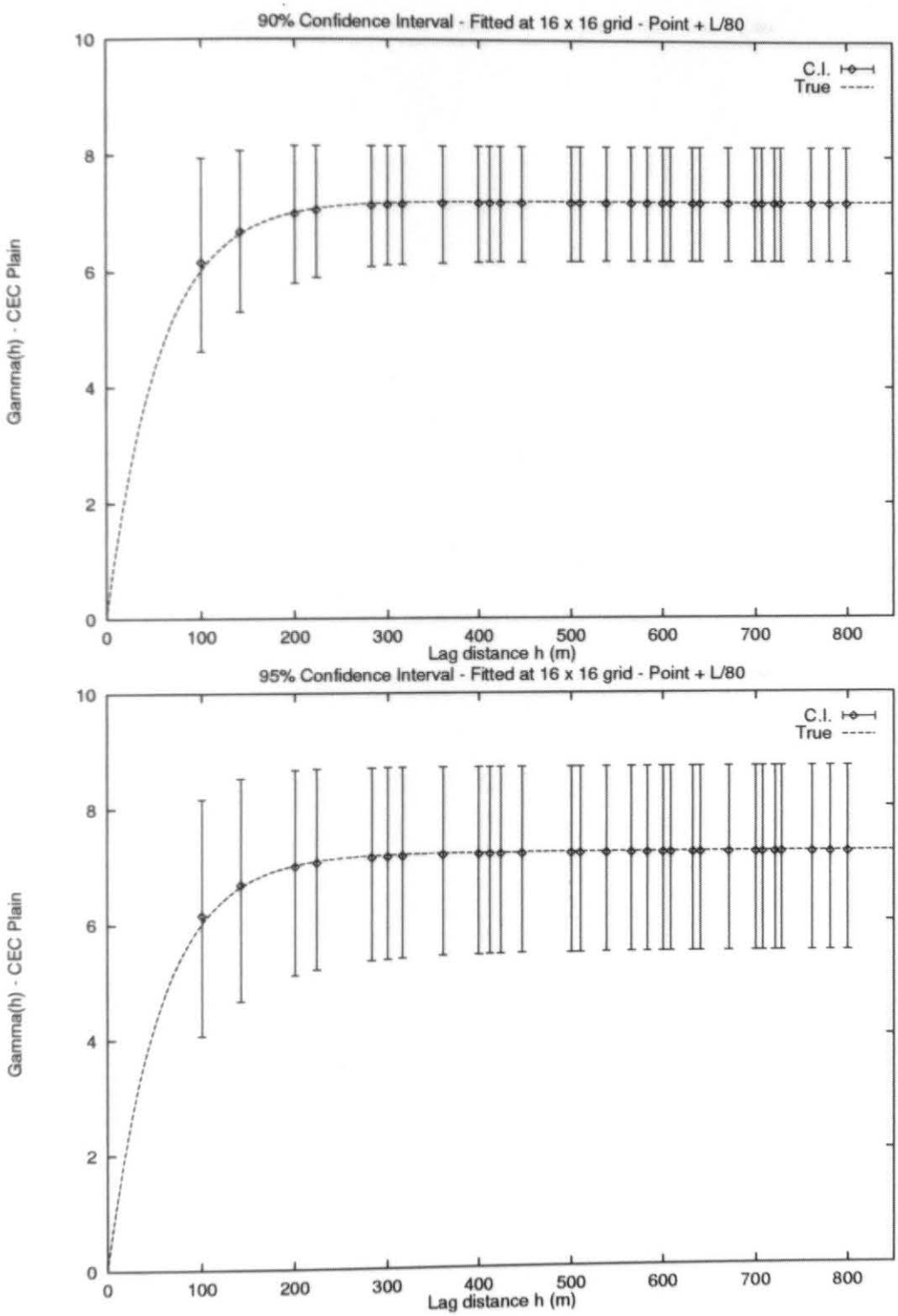


Figure B.14: Numerical Experiment 1 Plain, 16x16 - point + L/80

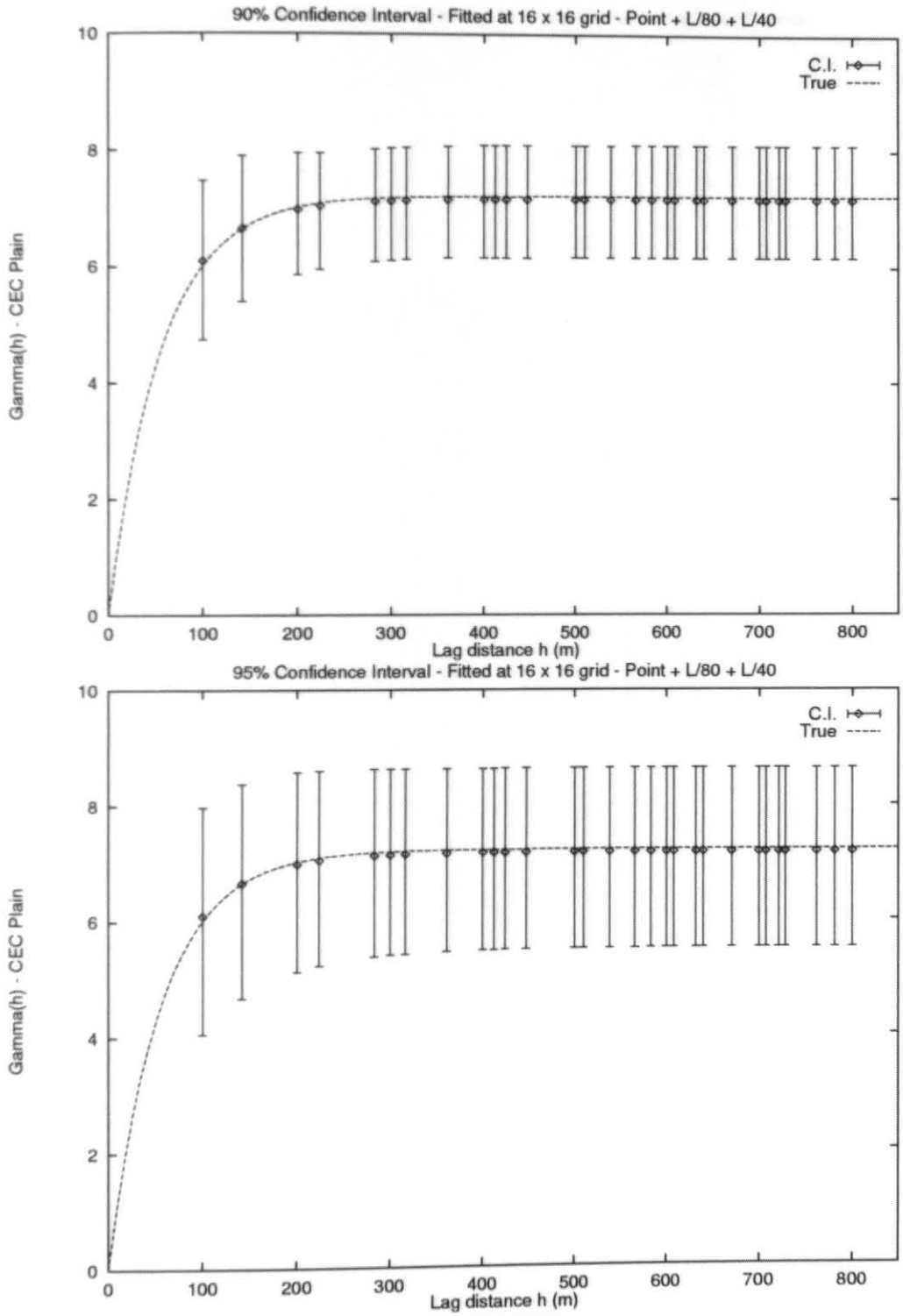


Figure B.15: Numerical Experiment 1 Plain, 16x16 - point + L/80 + L/40

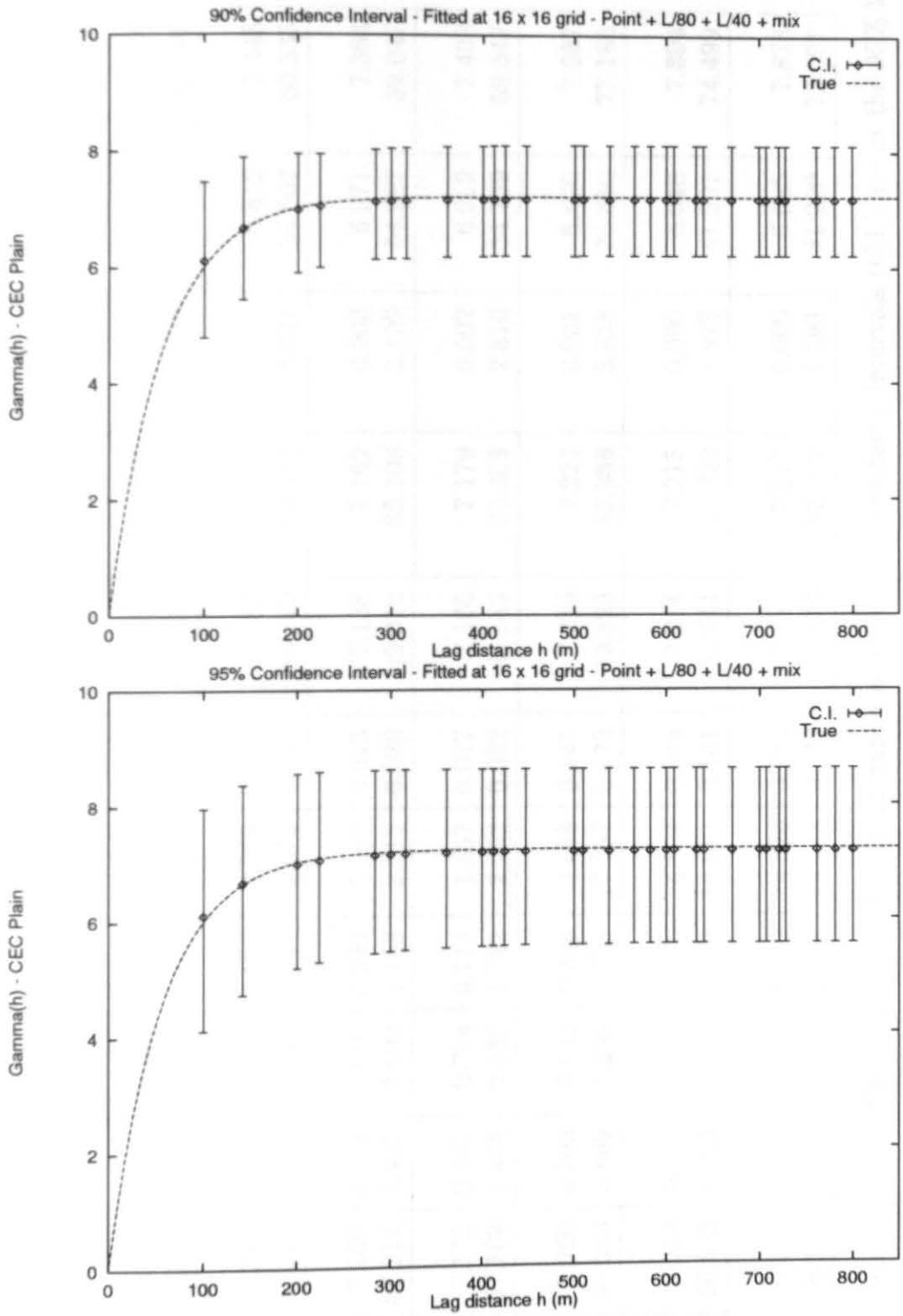


Figure B.16: Numerical Experiment 1 Plain, 16x16 - point + L/80 + L/40 + mix

Parameter	Mean	Bias	%Bias	S.D.	%S.D.	C.V.	Low C.I.	High C.I.	S.D. C.I.	Low C.I.	High C.I.
Point:											
Sill	7.195	0.028	0.383	0.425	5.906	0.059	7.195	7.196	0.001	6.439	8.045
Range	55.343	0.378	0.682	4.132	7.466	0.075	47.154	63.533	7.289	46.557	62.395
L/80:											
Sill	7.186	0.037	0.512	0.163	2.269	0.023	7.182	7.190	0.002	6.812	7.448
Range	57.335	1.614	2.814	1.534	2.676	0.027	48.279	66.391	3.931	54.607	60.335
L/40:											
Sill	7.150	0.073	1.016	0.090	1.260	0.013	7.138	7.162	0.003	6.971	7.364
Range	57.214	1.493	2.609	1.153	2.015	0.020	49.319	65.108	2.109	54.987	59.047
Mixed:											
Sill	7.172	0.051	0.704	0.119	1.657	0.017	7.166	7.179	0.002	6.919	7.408
Range	57.169	1.448	2.532	1.282	2.242	0.022	48.945	65.392	2.816	54.759	59.547
Case A:											
Sill	7.220	0.003	0.042	0.338	4.688	0.047	7.219	7.221	0.001	6.630	7.957
Range	56.590	0.869	1.536	9.788	17.297	0.173	53.823	59.358	5.818	35.880	77.186
Case B:											
Sill	7.214	0.009	0.119	0.325	4.504	0.045	7.214	7.215	0.000	6.648	7.894
Range	56.683	0.962	1.697	7.976	14.071	0.141	55.644	57.722	1.973	41.291	74.499
Case C:											
Sill	7.217	0.006	0.089	0.312	4.326	0.043	7.216	7.217	0.000	6.685	7.878
Range	56.670	0.949	1.675	7.887	13.918	0.139	55.873	57.467	1.393	41.220	73.997

Table B.3: Experiment 1, Plain, 24 x 24, sill = 7.223, range = 3*55.721. Confidence Intervals (C.I.) are at the 95% level

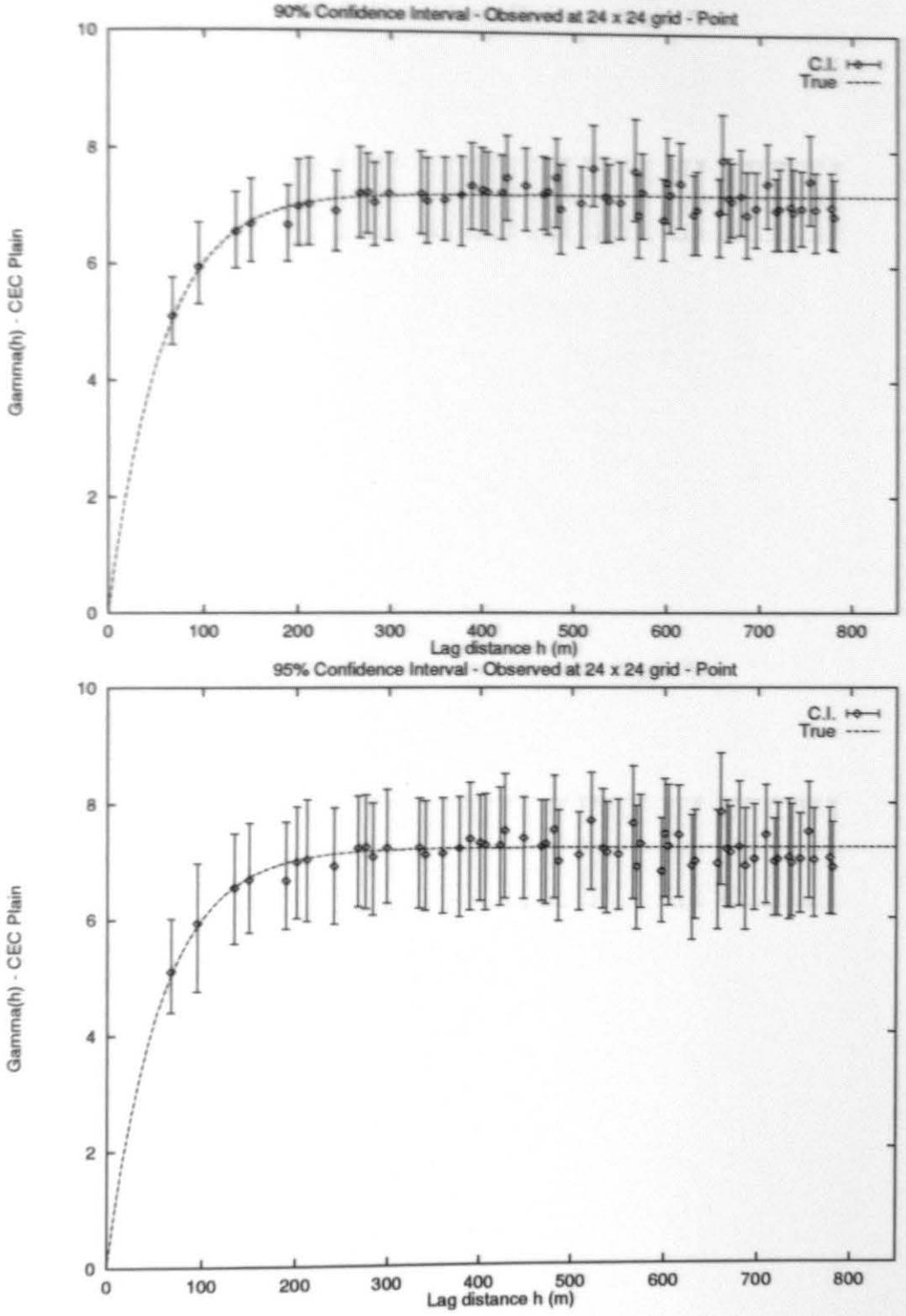


Figure B.17: Numerical Experiment 1 Plain, 24x24 - Obs. Point

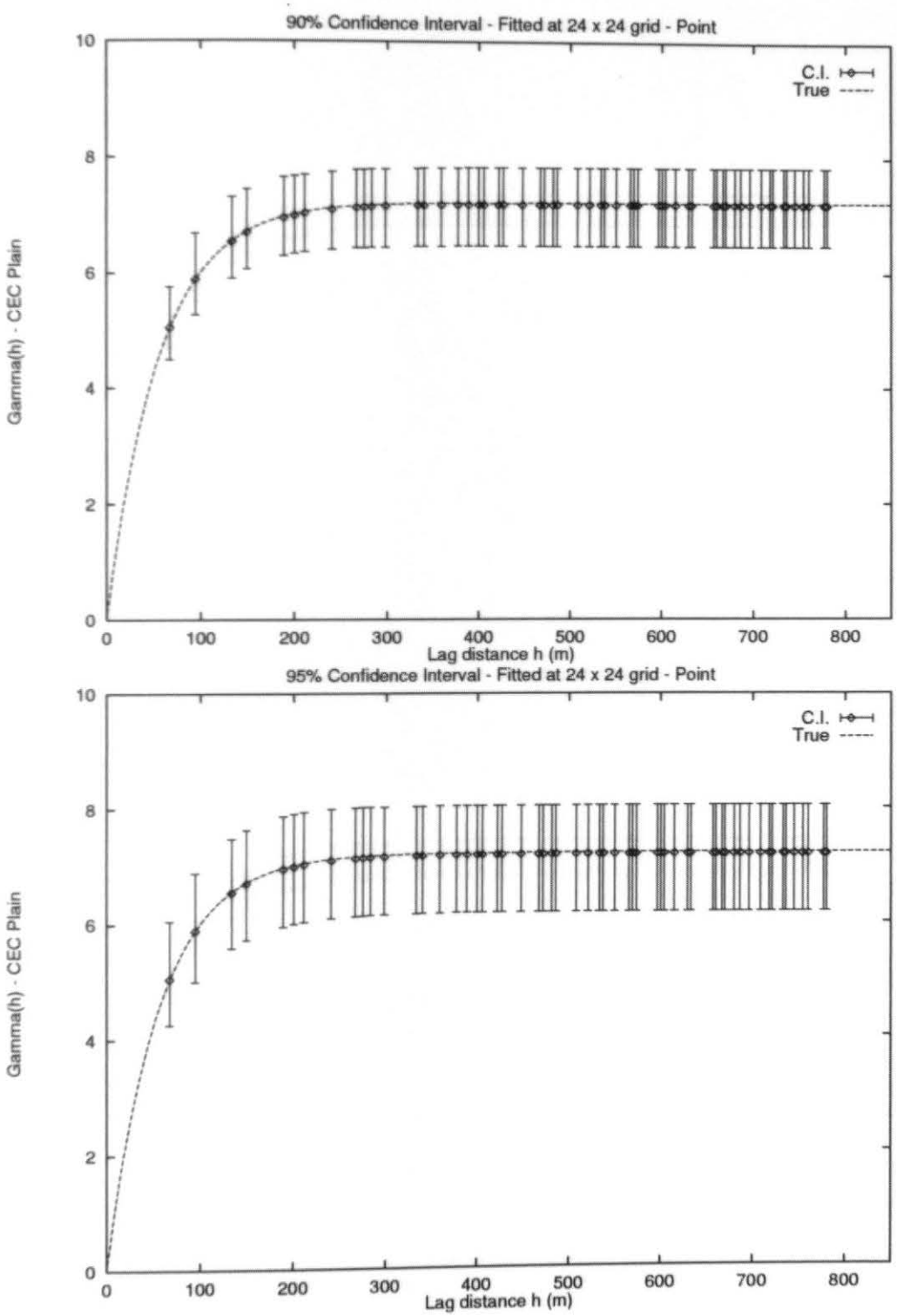


Figure B.18: Numerical Experiment 1 Plain, 24x24 - Fit. Point

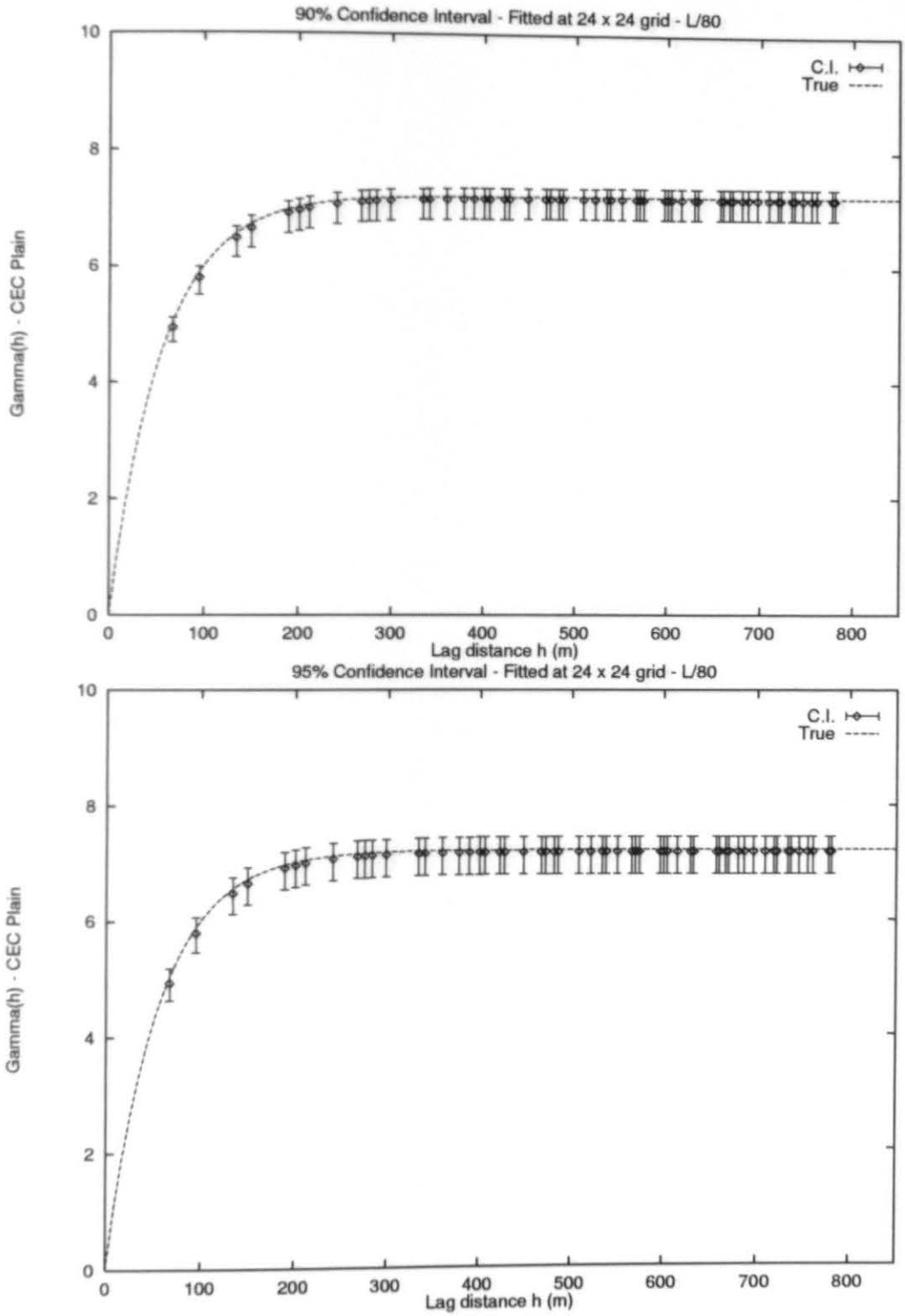


Figure B.19: Numerical Experiment 1 Plain, 24x24 - L/80

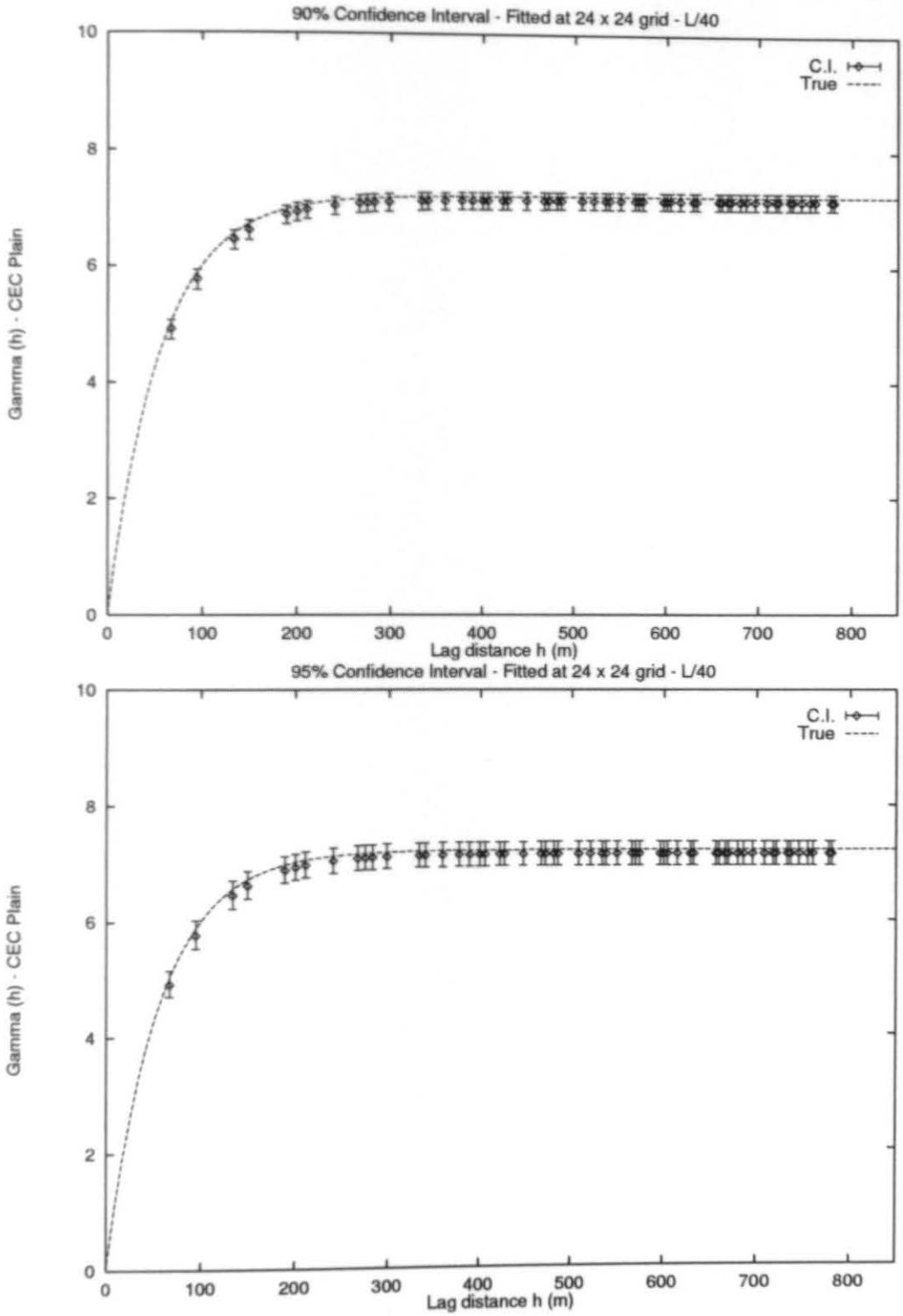


Figure B.20: Numerical Experiment 1 Plain, 24x24 - L/40

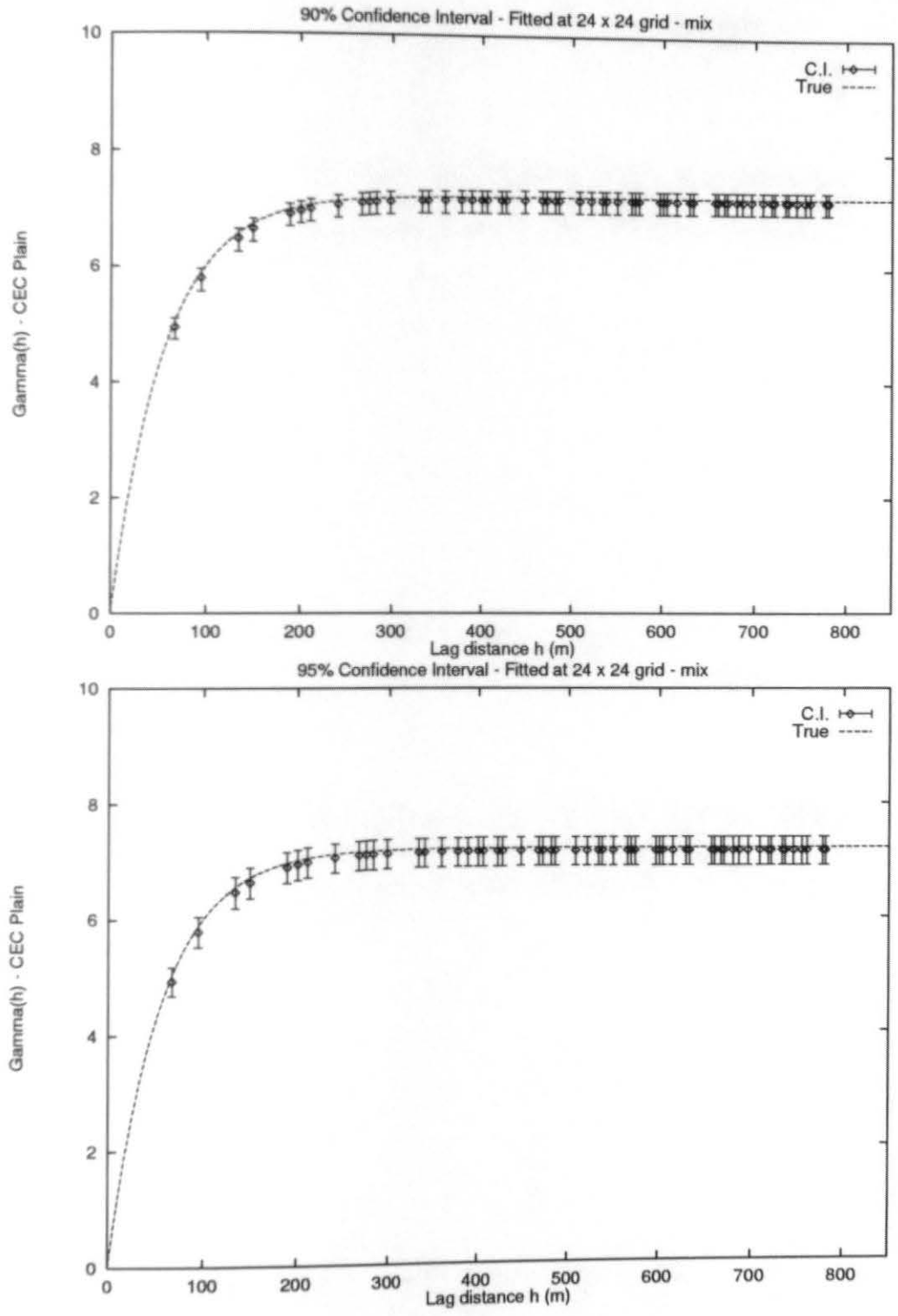


Figure B.21: Numerical Experiment 1 Plain, 24x24 - mix

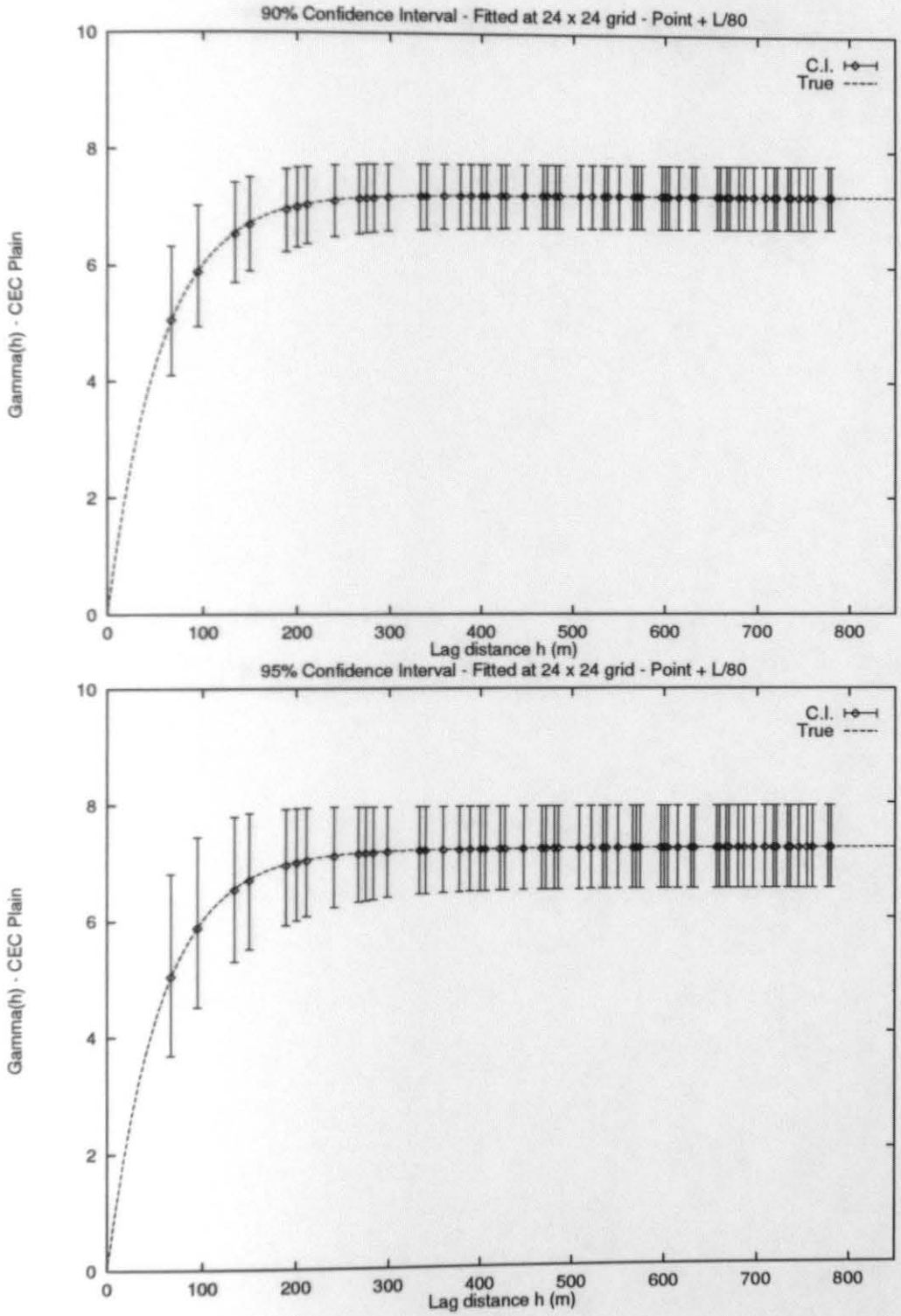


Figure B.22: Numerical Experiment 1 Plain, 24x24 - point + L/80

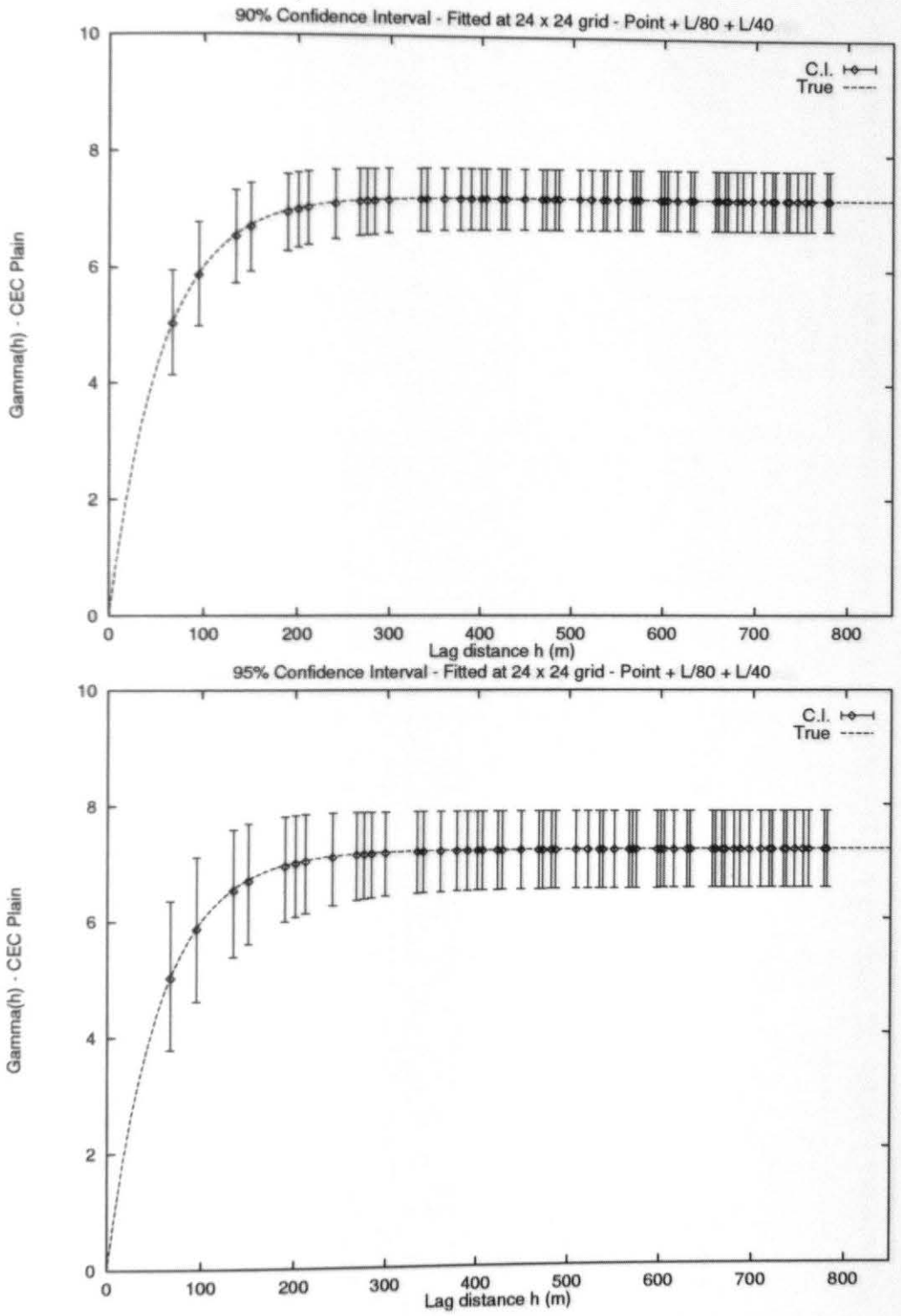


Figure B.23: Numerical Experiment 1 Plain, 24x24 - point + L/80 + L/40

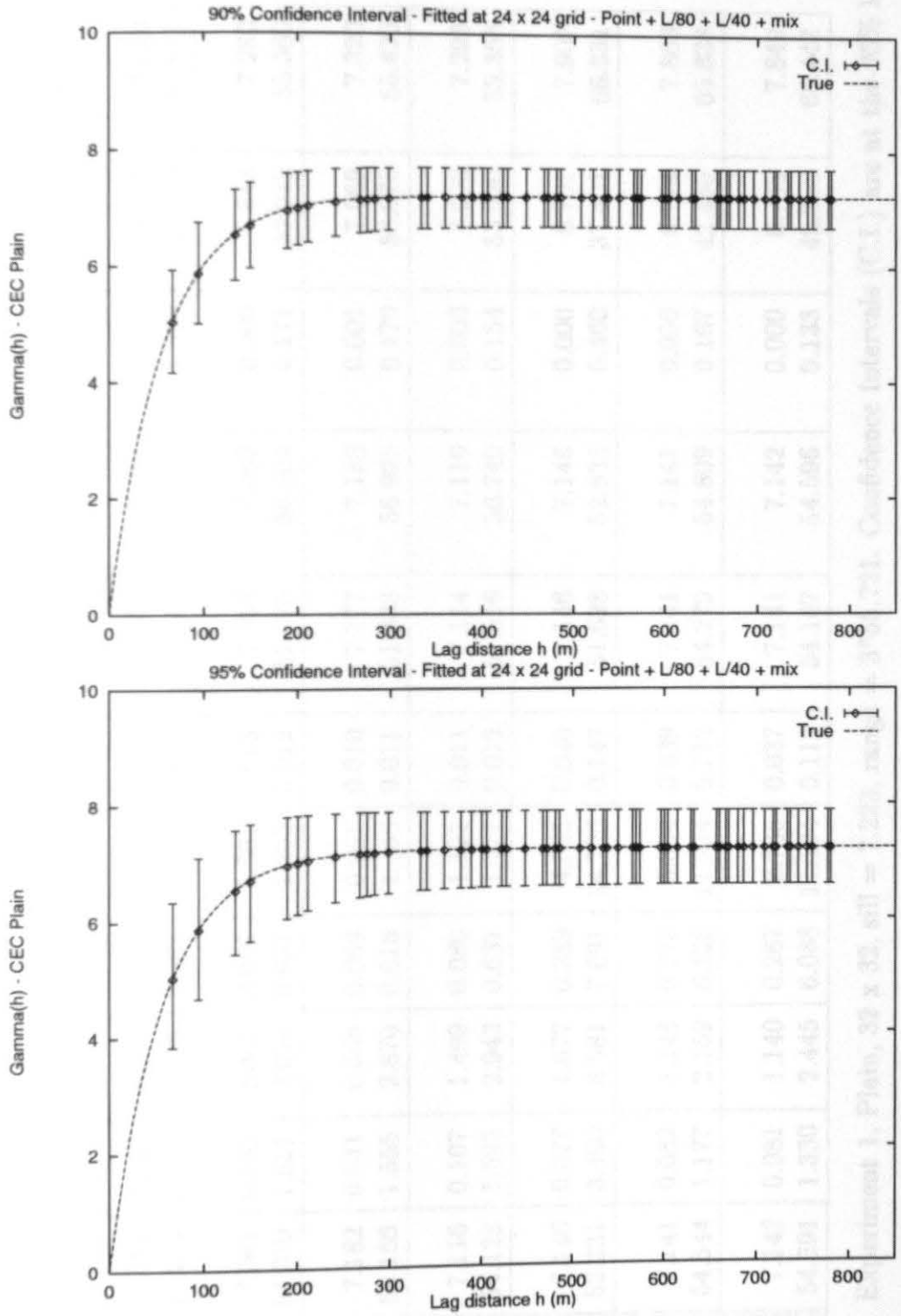


Figure B.24: Numerical Experiment 1 Plain, 24x24 - point + L/80 + L/40 + mix

Parameter	Mean	Bias	%Bias	S.D.	%S.D.	C.V.	Low C.I.	High C.I.	S.D. C.I.	Low C.I.	High C.I.
Point:											
Sill	7.161	0.062	0.872	0.346	4.827	0.048	7.160	7.161	0.000	6.603	8.010
Range	55.204	0.517	0.936	2.682	4.858	0.049	52.314	58.095	1.578	50.283	60.675
L/80:											
Sill	7.061	0.162	2.301	0.092	1.307	0.013	7.059	7.062	0.000	6.943	7.280
Range	54.070	1.651	3.054	0.651	1.203	0.012	51.576	56.564	0.171	52.912	55.360
L/40:											
Sill	7.182	0.041	0.568	0.069	0.961	0.010	7.177	7.188	0.001	7.069	7.320
Range	54.166	1.555	2.870	0.616	1.136	0.011	51.338	56.995	0.170	53.075	55.426
Mixed:											
Sill	7.116	0.107	1.499	0.080	1.122	0.011	7.114	7.119	0.000	7.000	7.298
Range	54.128	1.593	2.943	0.637	1.177	0.012	51.516	56.740	0.154	53.020	55.368
Case A:											
Sill	7.146	0.077	1.077	0.289	4.045	0.040	7.146	7.146	0.000	6.701	7.907
Range	52.231	3.490	6.681	7.691	14.726	0.147	51.628	52.835	0.460	37.401	66.331
Case B:											
Sill	7.141	0.082	1.148	0.275	3.854	0.039	7.141	7.141	0.000	6.714	7.863
Range	54.544	1.177	2.159	6.128	11.234	0.112	54.279	54.809	0.167	42.455	65.838
Case C:											
Sill	7.142	0.081	1.140	0.267	3.735	0.037	7.141	7.142	0.000	6.735	7.849
Range	54.391	1.330	2.445	6.088	11.194	0.112	54.187	54.596	0.133	42.353	65.442

Table B.4: Experiment 1, Plain, 32 x 32, sill = 7.223, range = 3*55.721. Confidence Intervals (C.I.) are at the 95% level

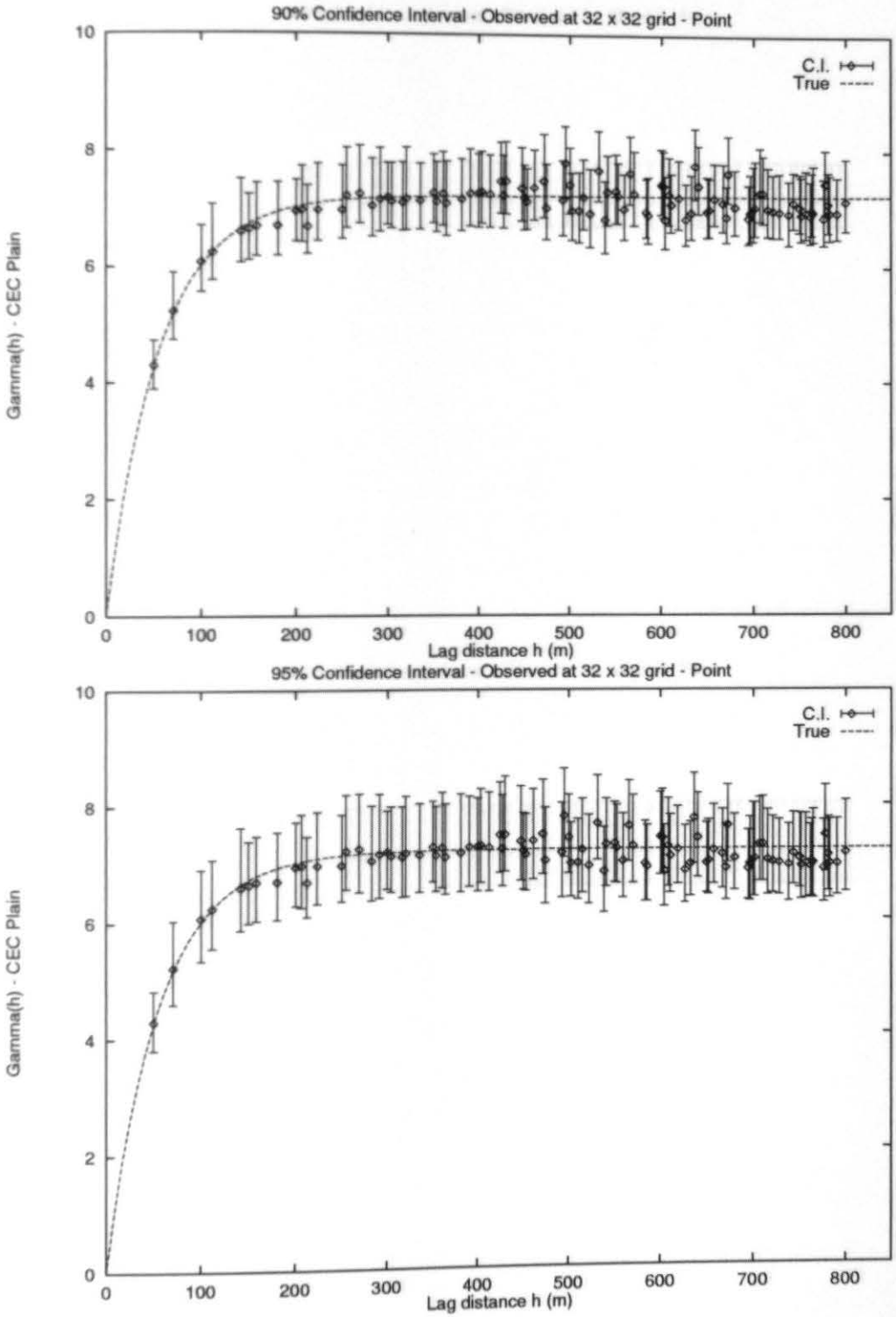


Figure B.25: Numerical Experiment 1 Plain, 32x32 - Obs. Point

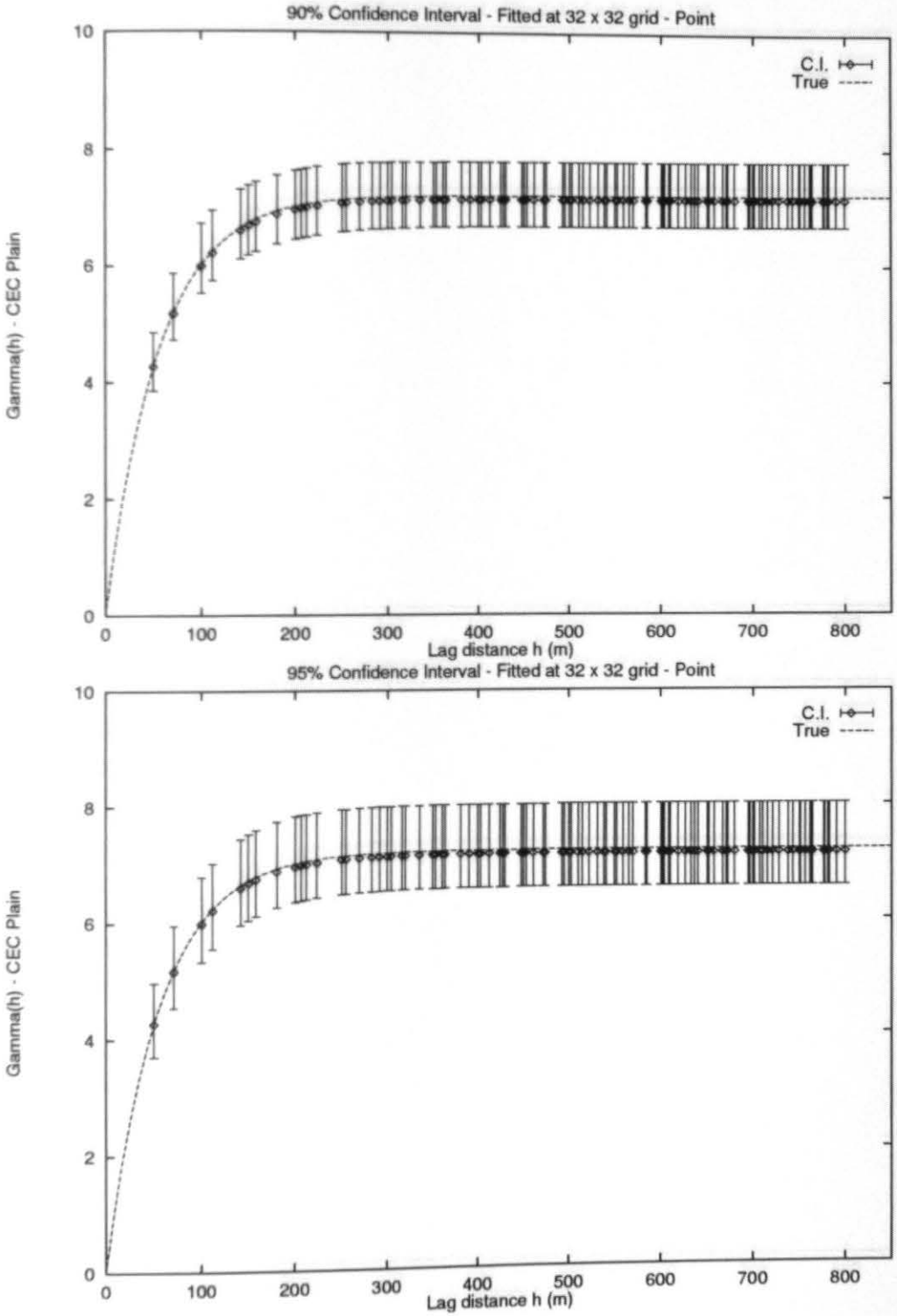


Figure B.26: Numerical Experiment 1 Plain, 32x32 - Fit. Point

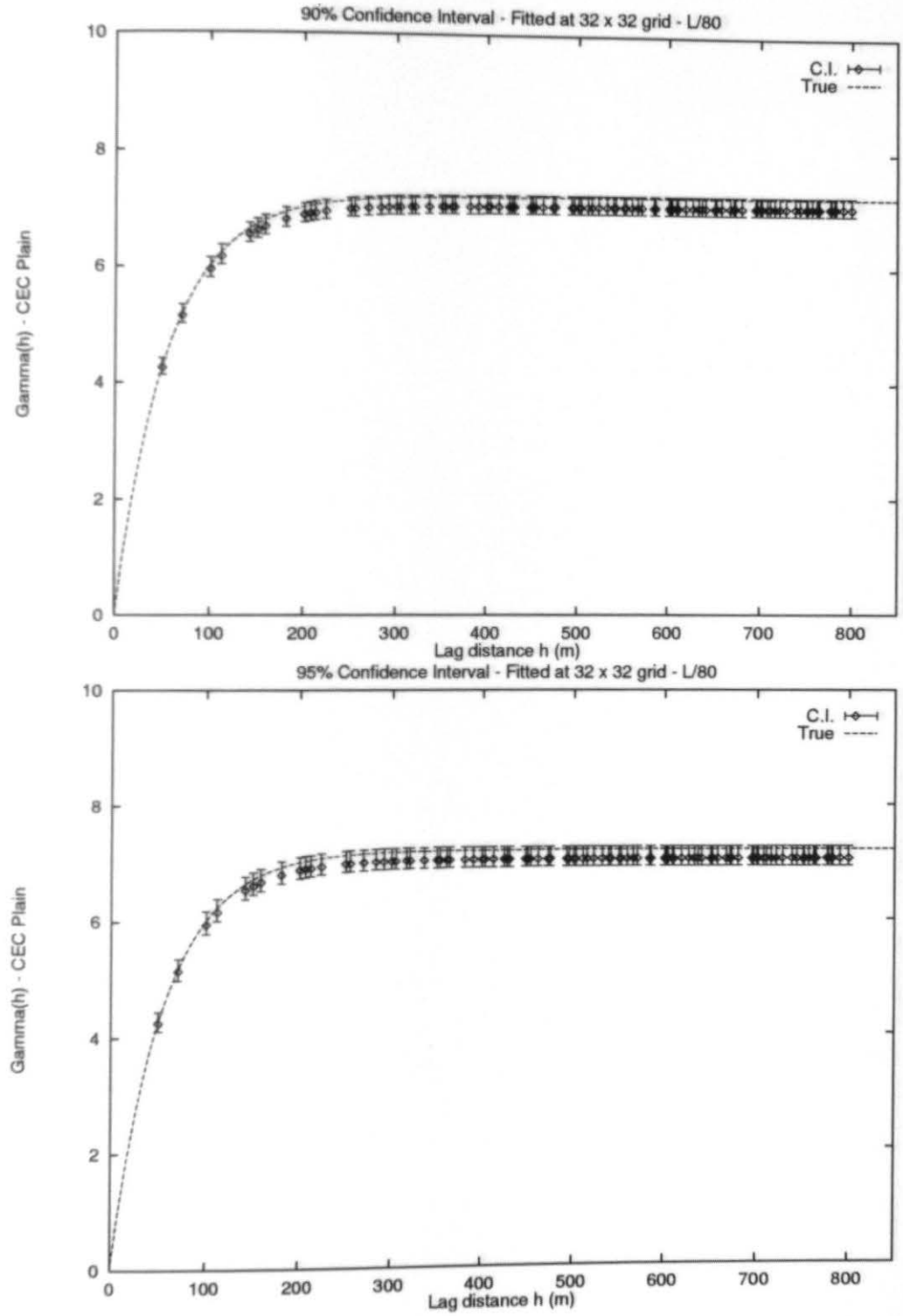


Figure B.27: Numerical Experiment 1 Plain, 32x32 - L/80

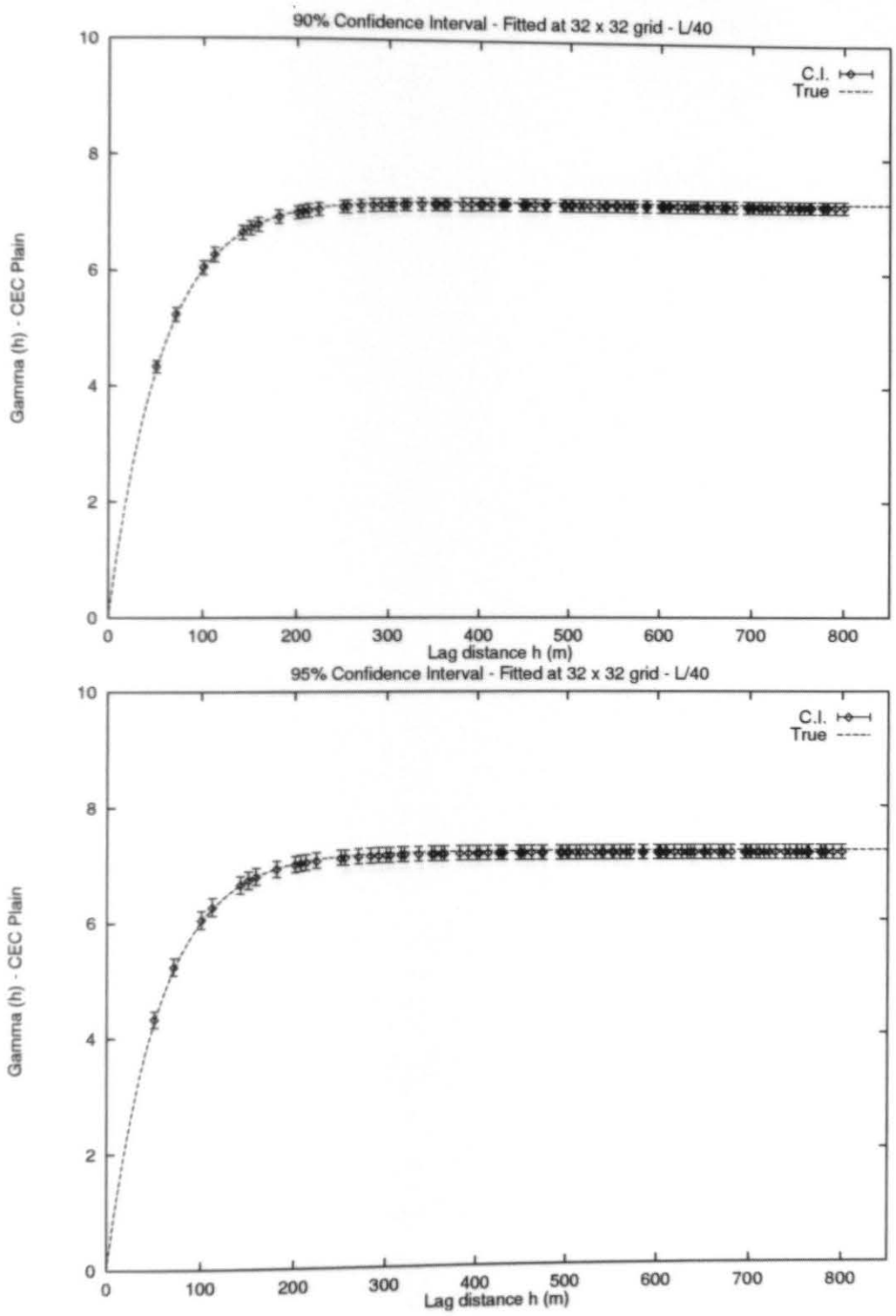


Figure B.28: Numerical Experiment 1 Plain, 32x32 - L/40

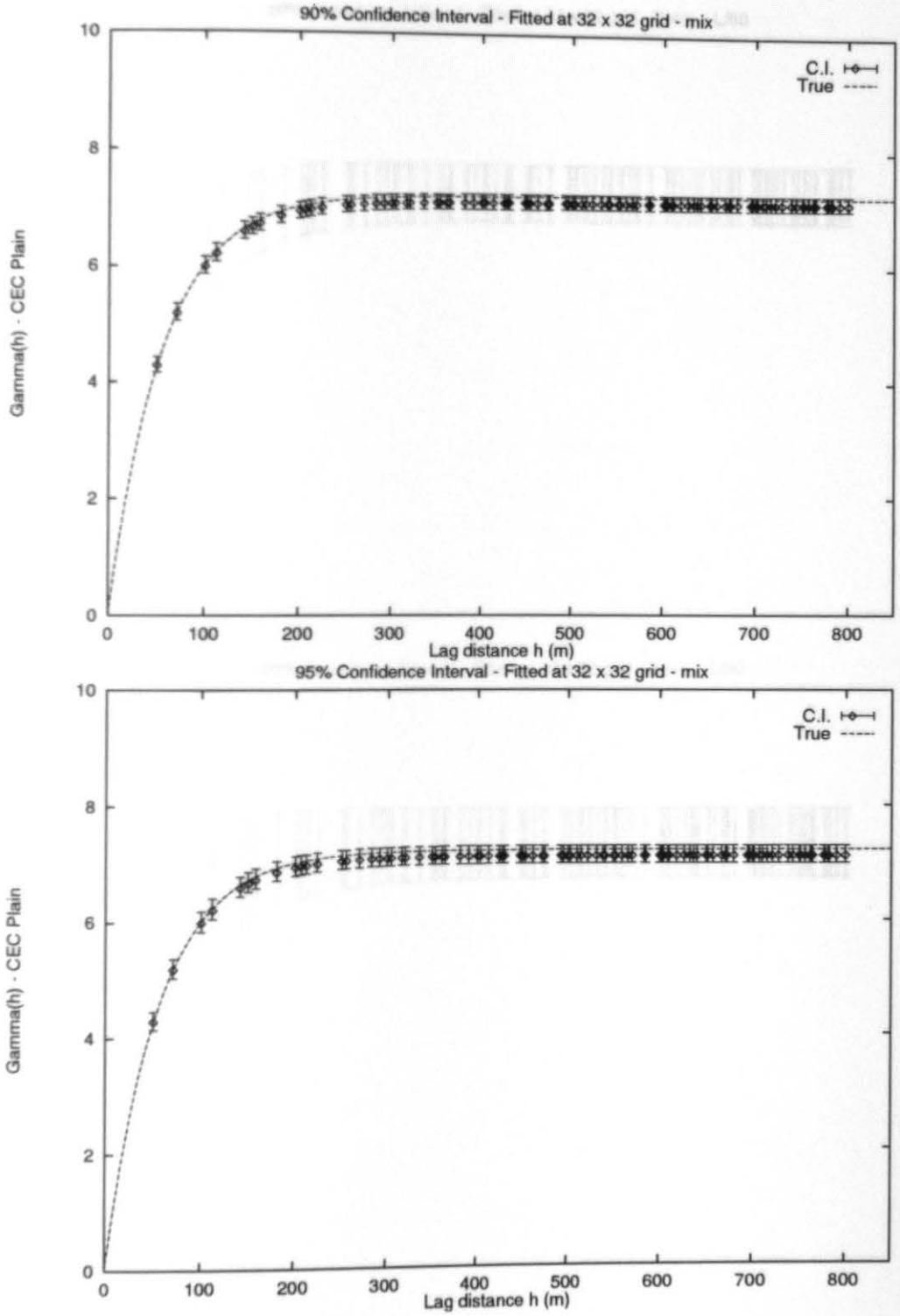


Figure B.29: Numerical Experiment 1 Plain, 32x32 - mix

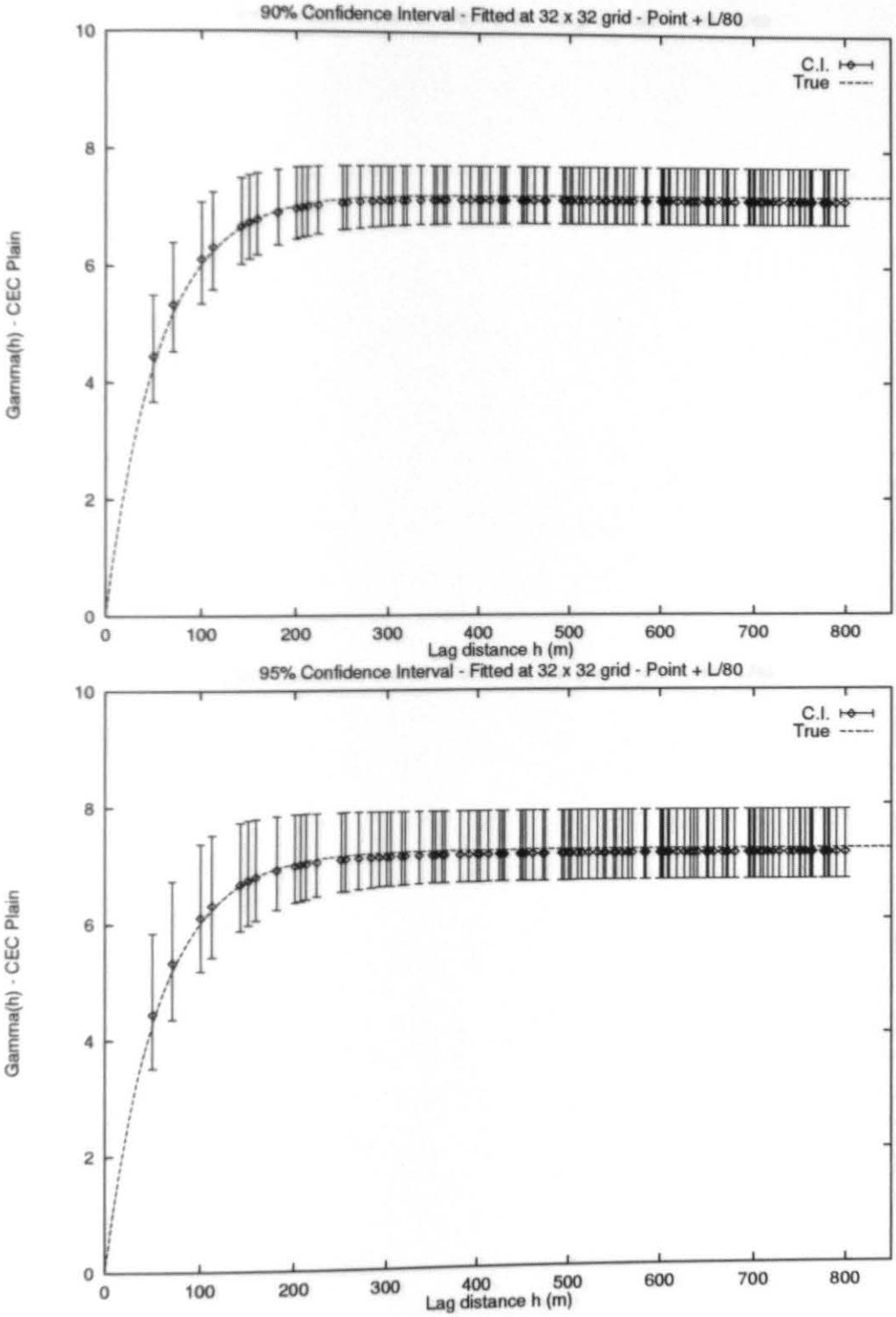


Figure B.30: Numerical Experiment 1 Plain, 32x32 - point + L/80

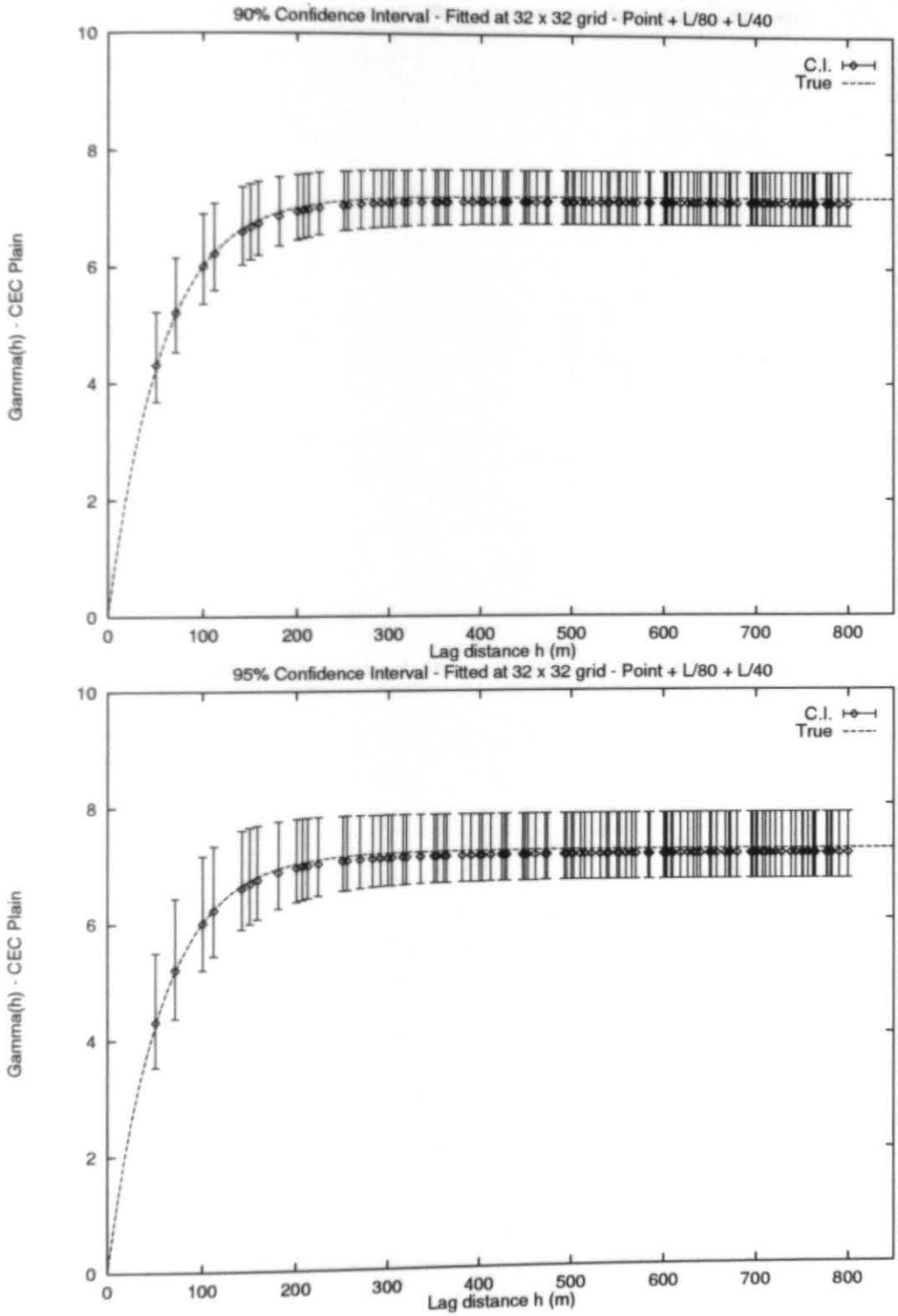


Figure B.31: Numerical Experiment 1 Plain, 32x32 - point + L/80 + L/40

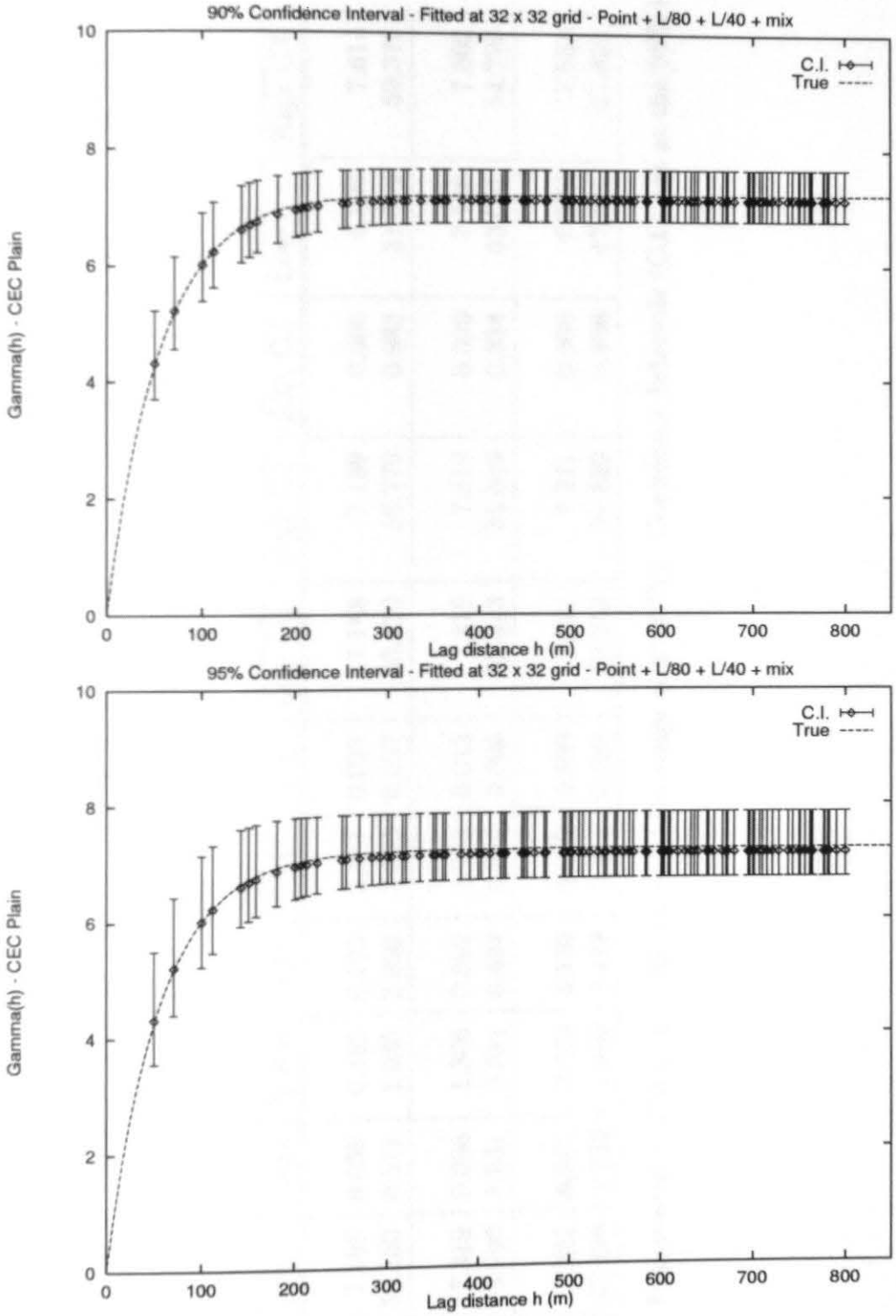


Figure B.32: Numerical Experiment 1 Plain, 32x32 - point + L/80 + L/40 + mix

Parameter	Mean	Bias	%Bias	S.D.	%S.D.	C.V.	Low C.I.	High C.I.	S.D. C.I.	Low C.I.	High C.I.
Point:											
Sill	7.188	0.035	0.482	0.193	2.684	0.027	7.188	7.189	0.000	6.909	7.614
Range	55.150	0.571	1.035	2.020	3.663	0.037	53.530	56.770	0.983	51.692	59.372
L/80:											
Sill	7.319	0.096	1.306	0.092	1.259	0.013	7.318	7.319	0.000	7.136	7.502
Range	53.790	1.931	3.591	0.494	0.919	0.009	52.523	55.056	0.334	53.051	54.720
Case A:											
Sill	7.221	0.002	0.023	0.146	2.022	0.020	7.221	7.221	0.000	6.997	7.553
Range	57.426	1.705	2.969	5.418	9.434	0.094	57.013	57.839	0.406	47.320	67.829

Table B.5: Experiment 1, Plain, 40 x 40, sill = 7.223, range = 3*55.721. Confidence Intervals (C.I.) are at the 95% level

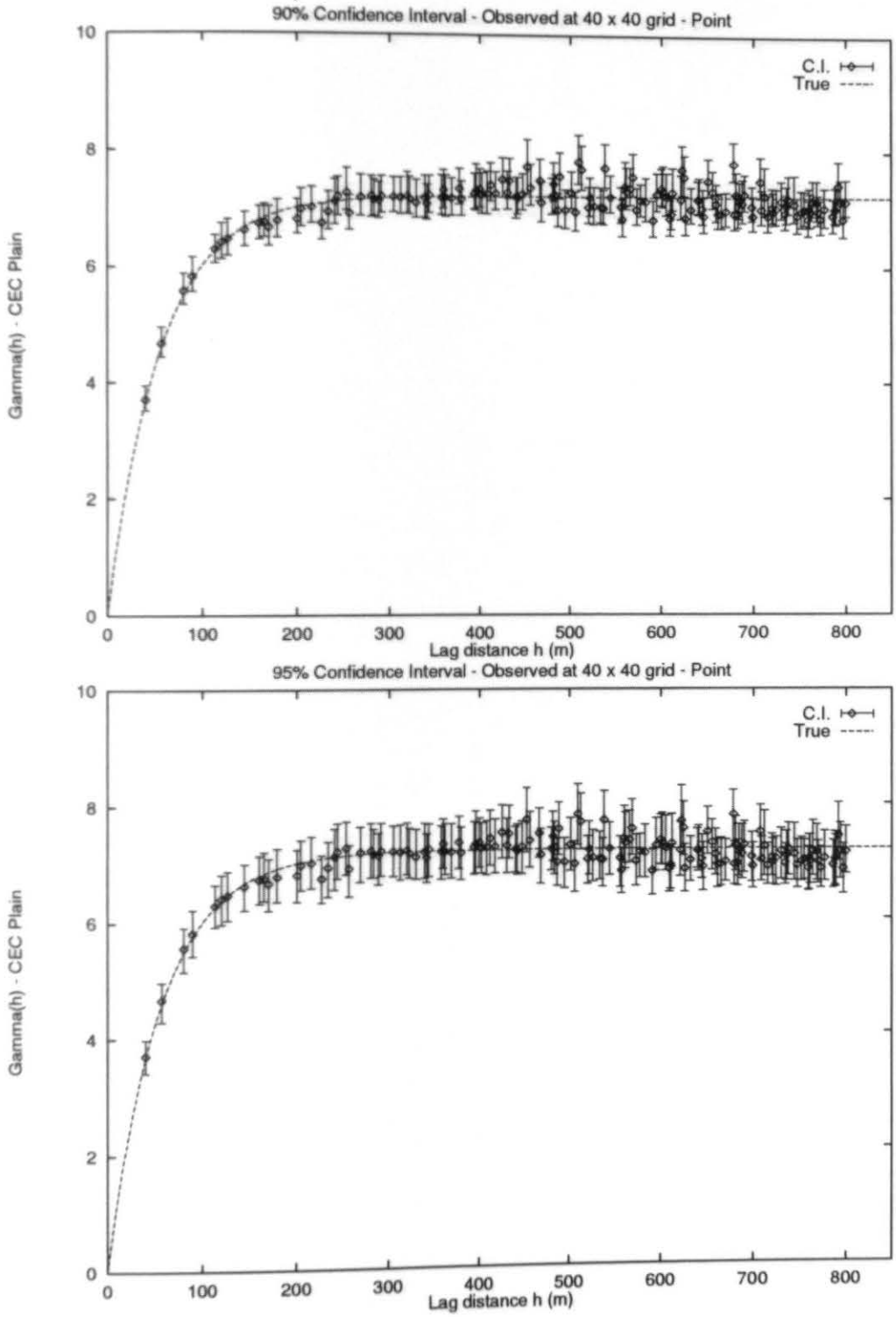


Figure B.33: Numerical Experiment 1 Plain, 40x40 - Obs. point

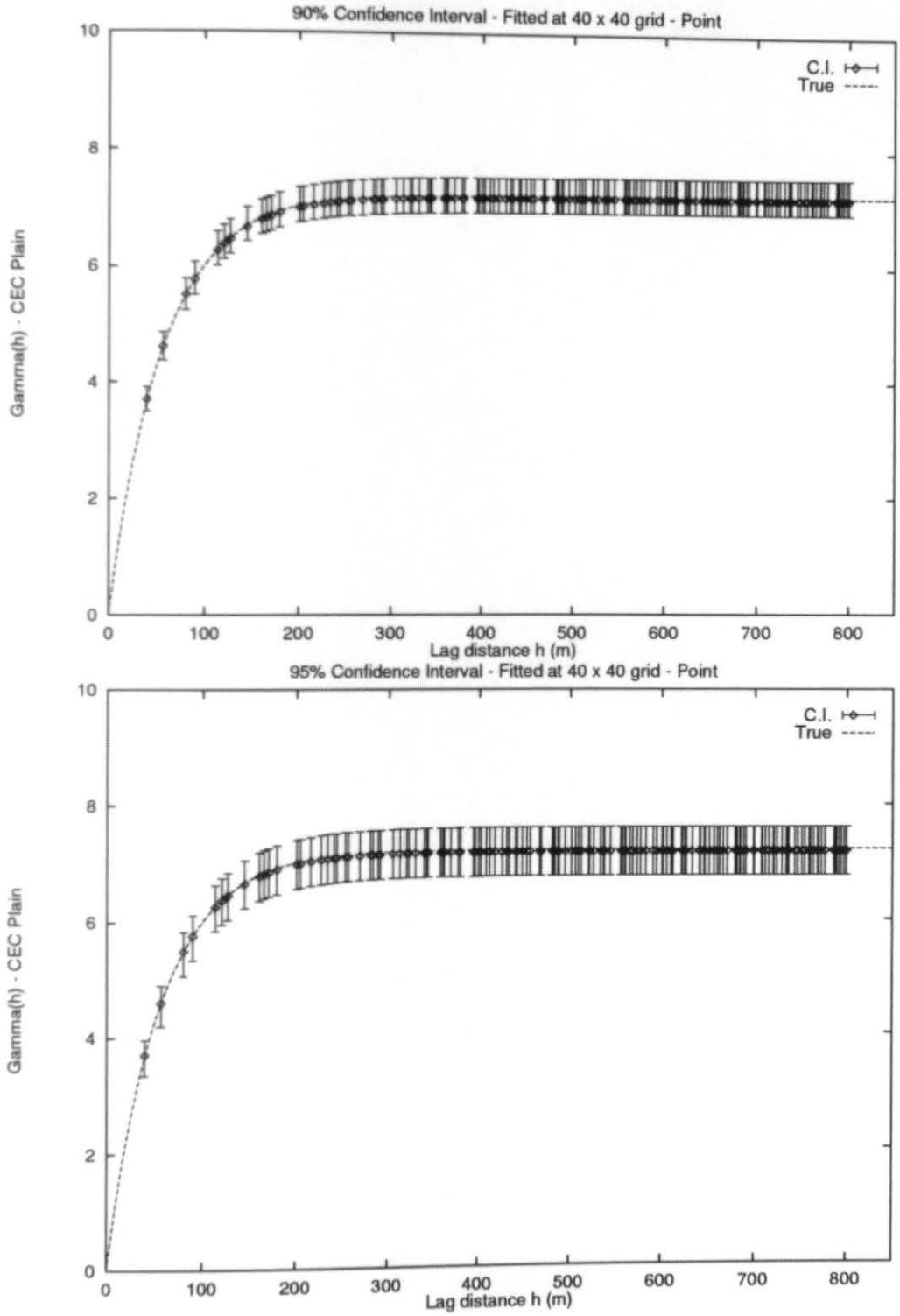


Figure B.34: Numerical Experiment 1 Plain, 40x40 - Fit. point

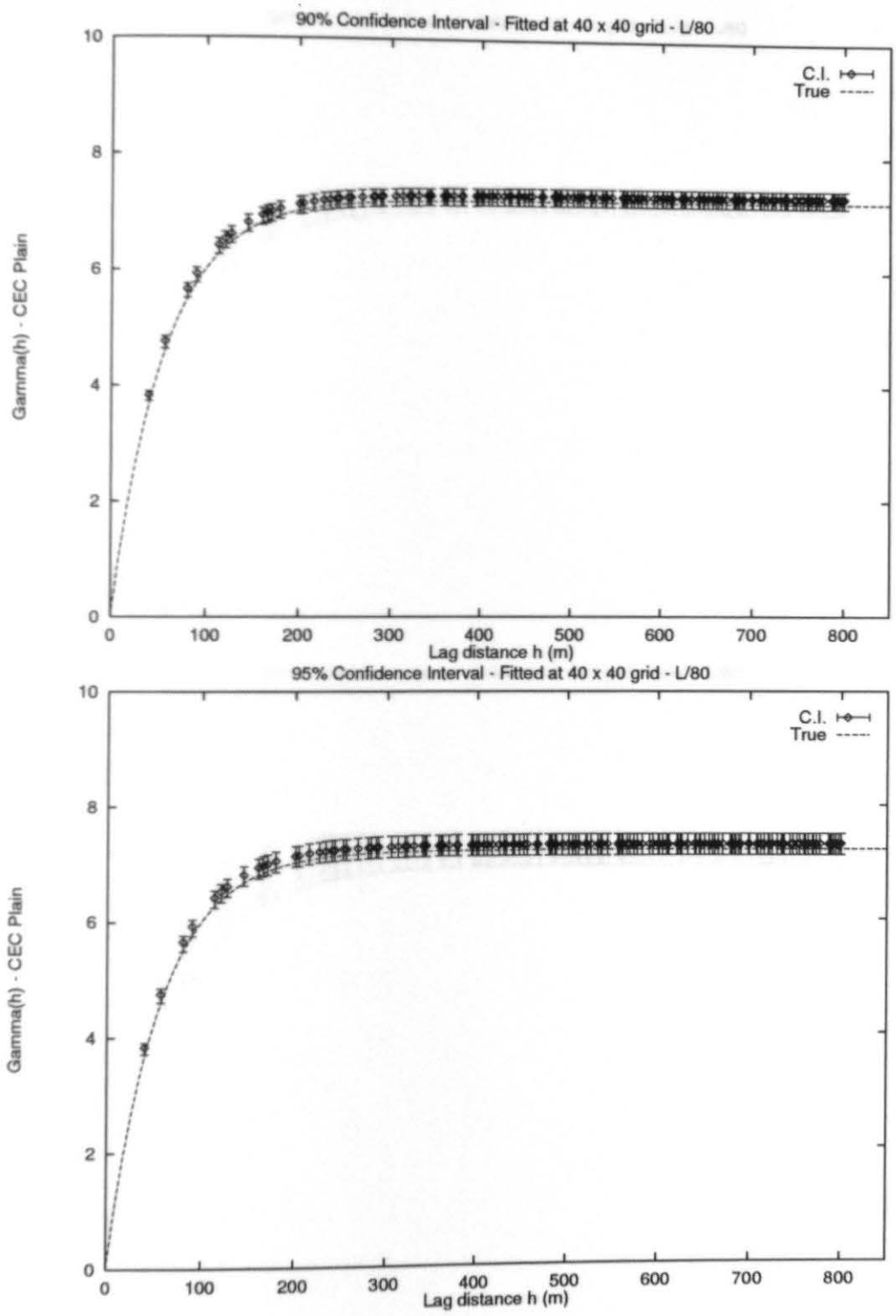


Figure B.35: Numerical Experiment 1 Plain, 40x40 - L/80

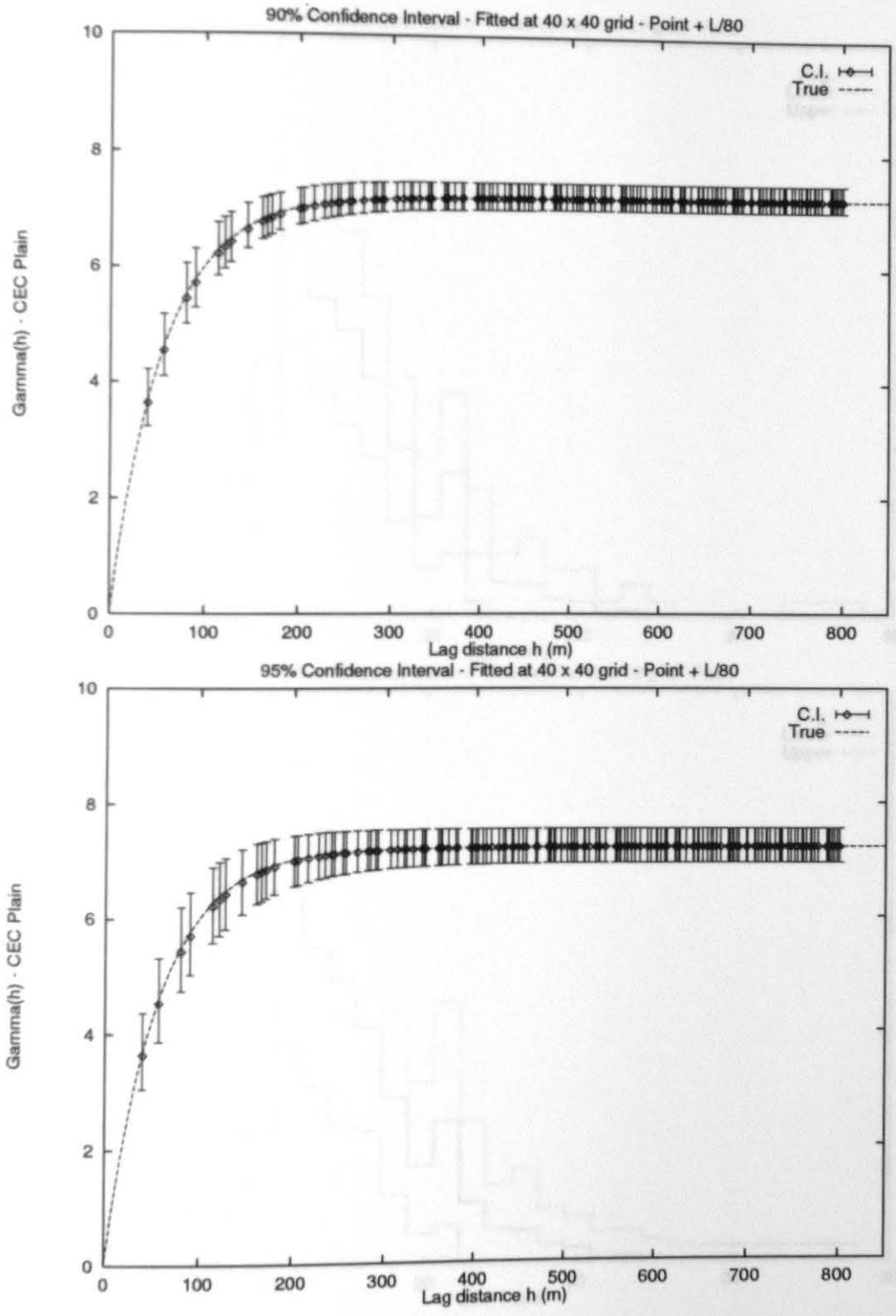


Figure B.36: Numerical Experiment 1 Plain, 40x40 - point + L/80

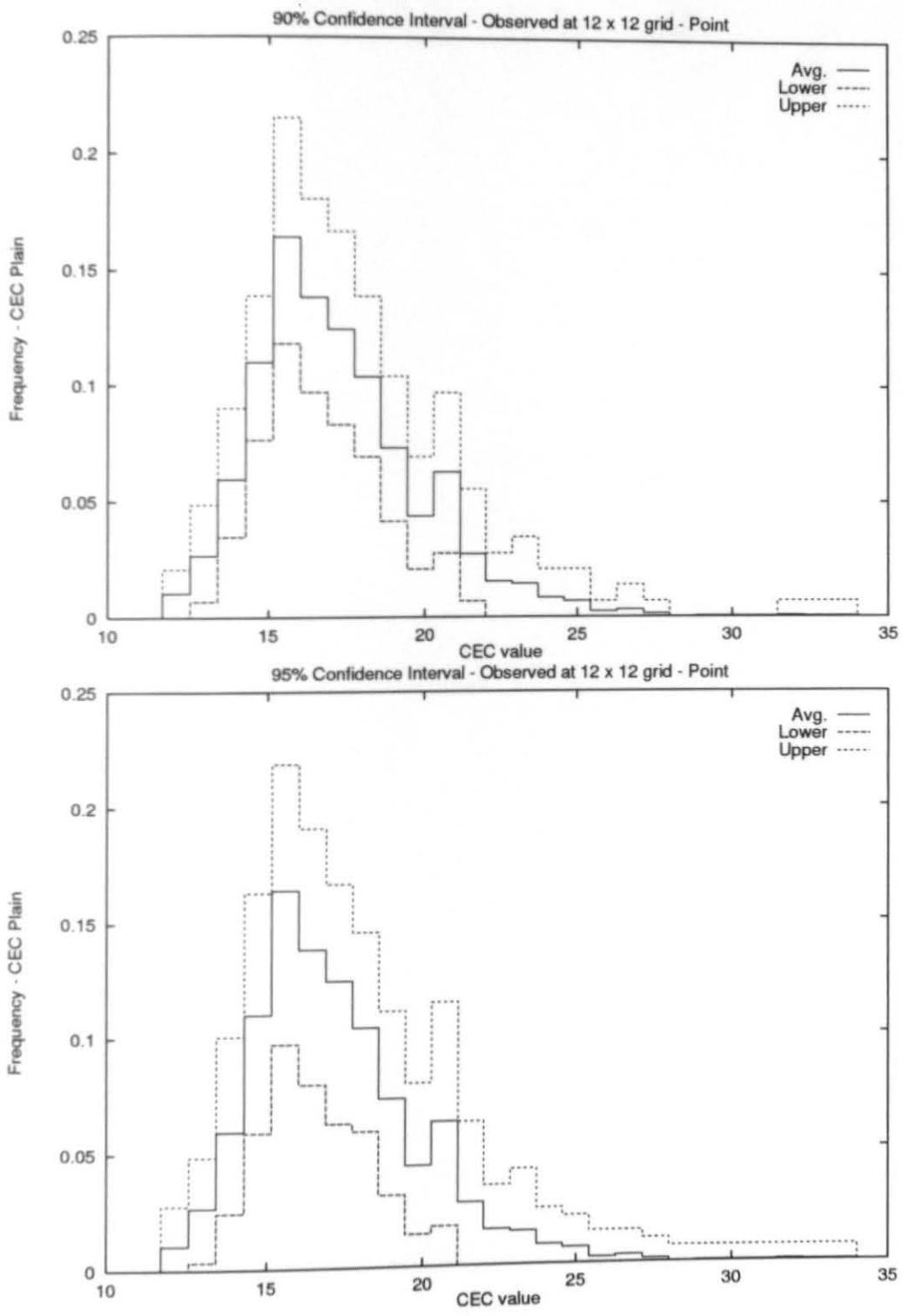


Figure B.37: Histogram Experiment 1 Plain, 12x12 - Obs. point

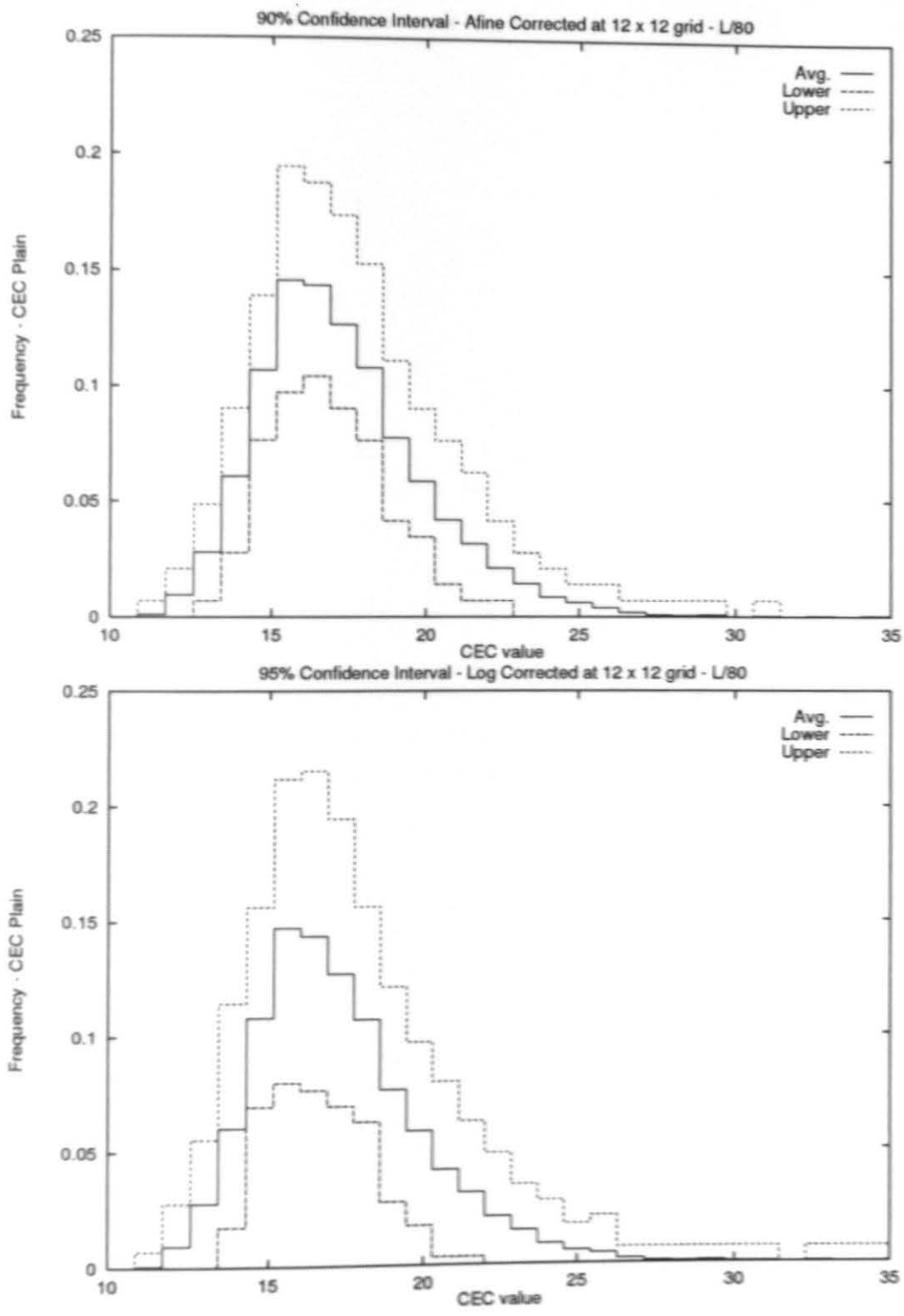


Figure B.38: Histogram Experiment 1 Plain, 12x12 - Afine L/80

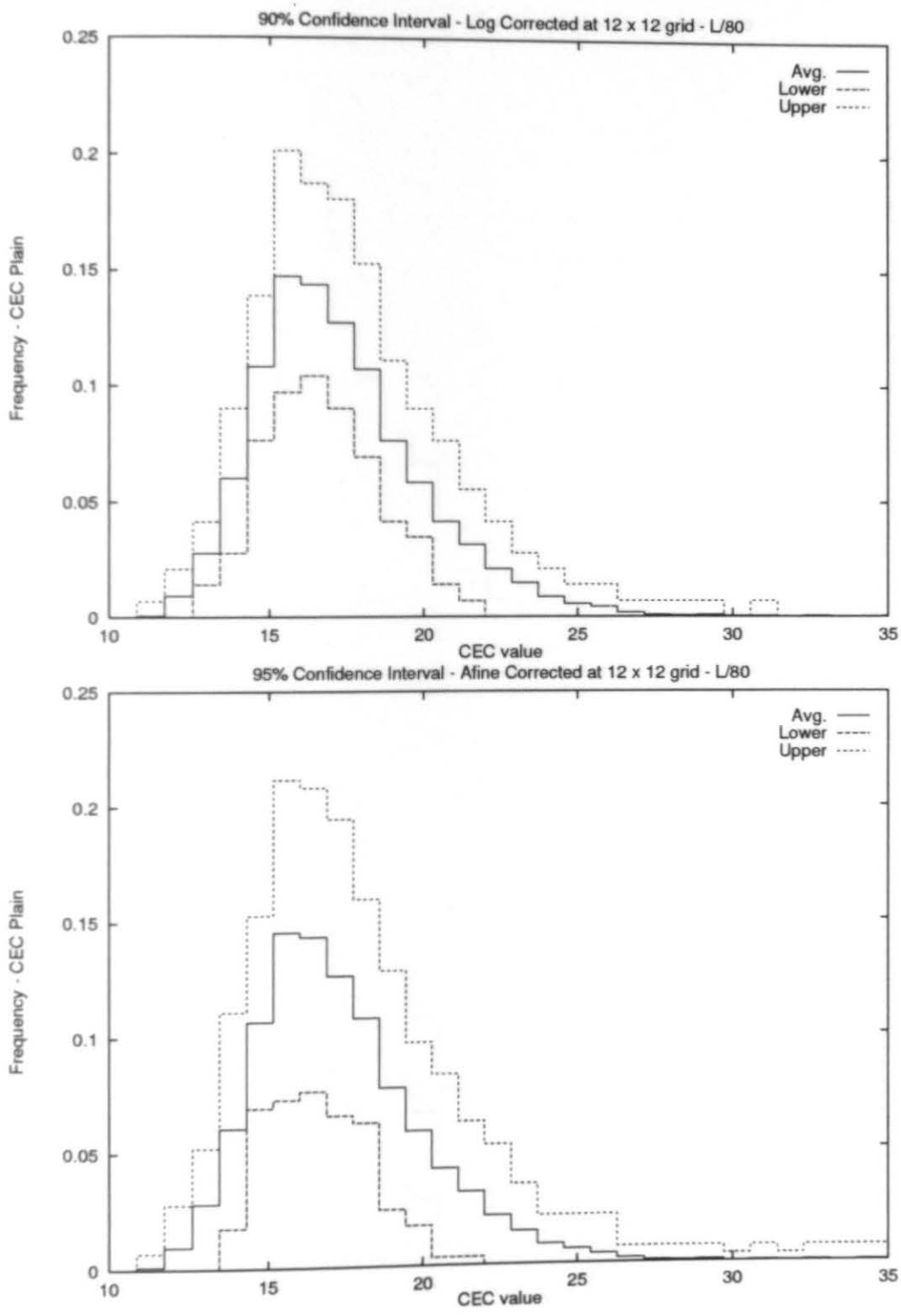


Figure B.39: Histogram Experiment 1 Plain, 12x12 - Log L/80

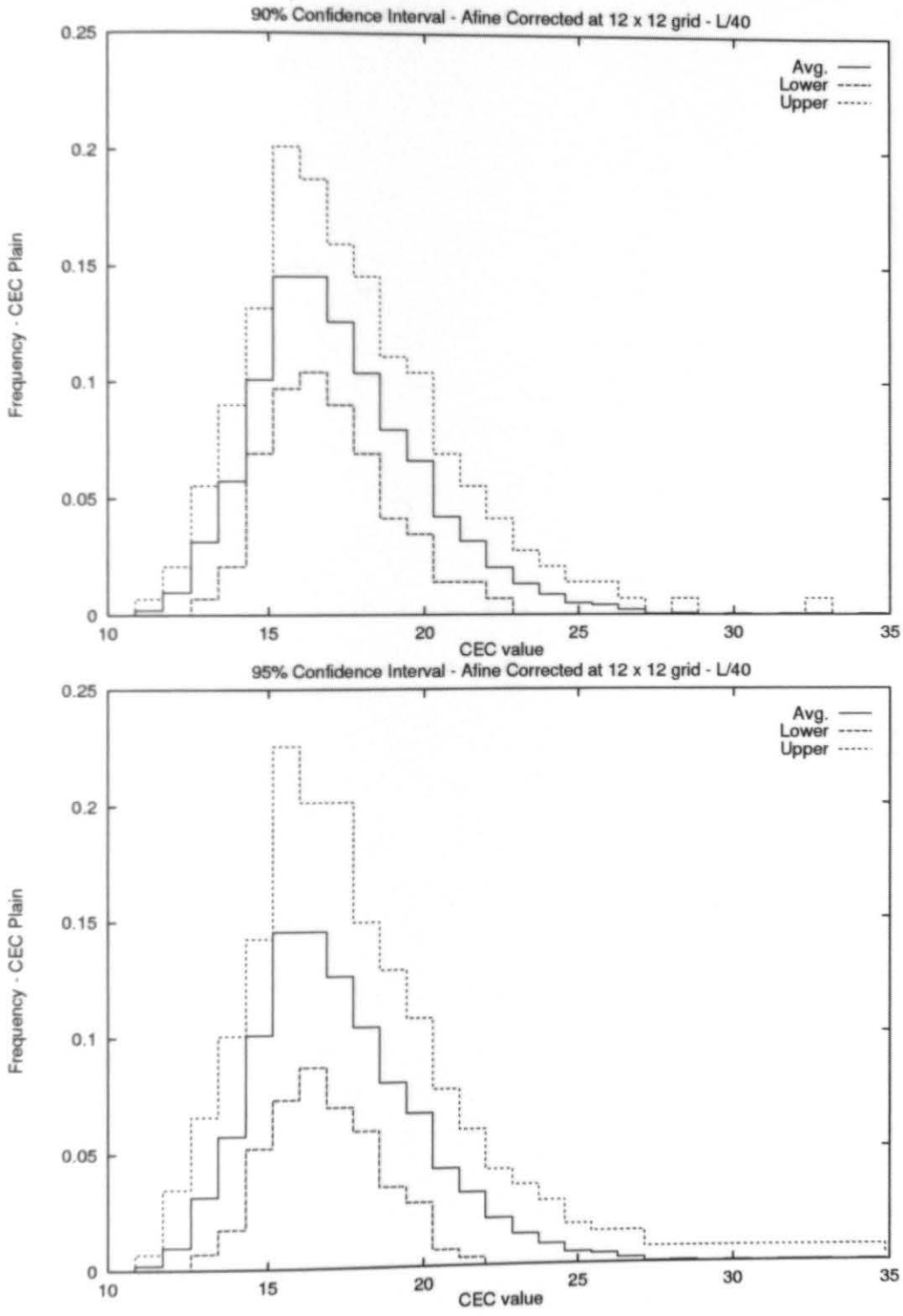


Figure B.40: Histogram Experiment 1 Plain, 12x12 - Afine L/40

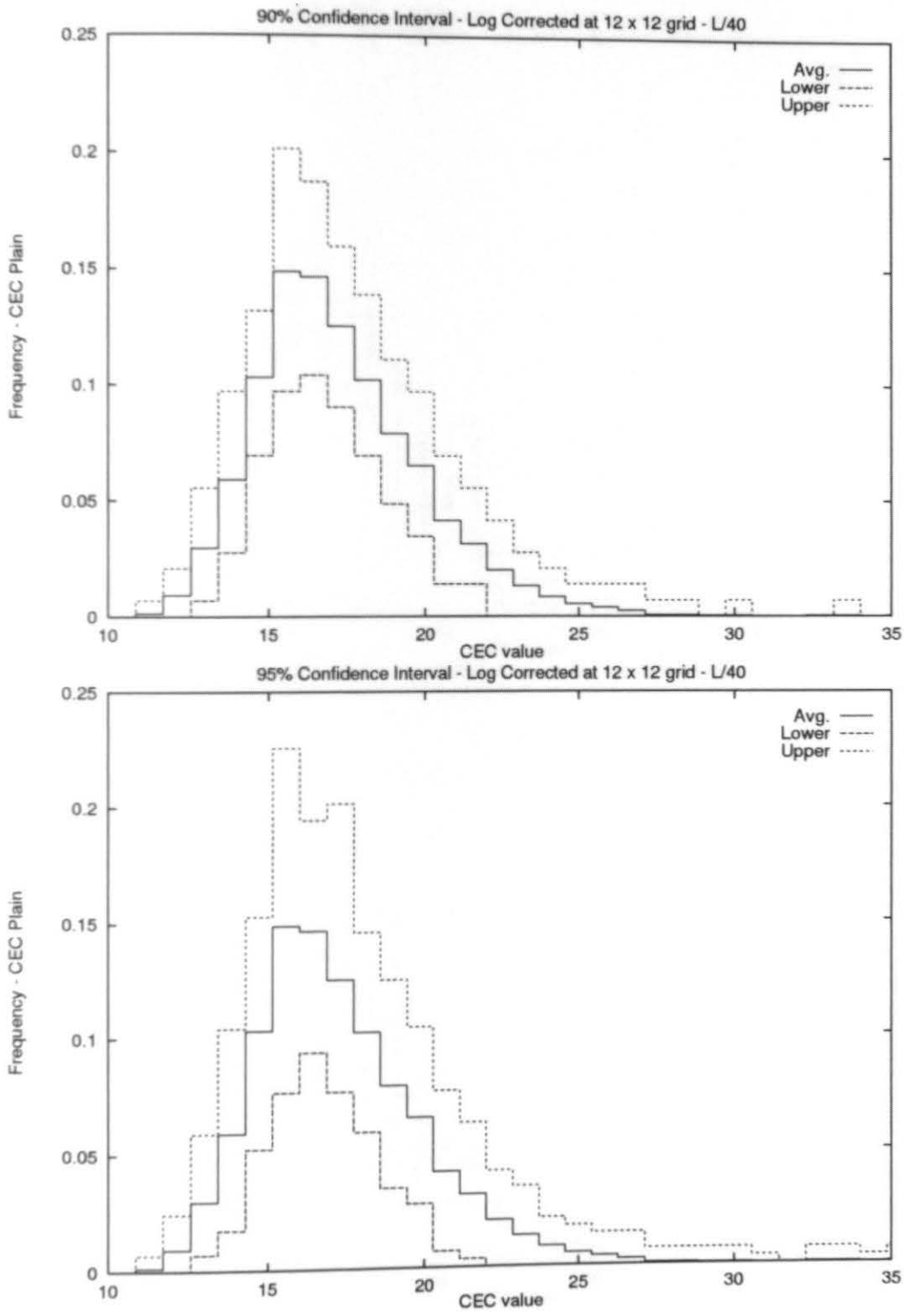


Figure B.41: Histogram Experiment 1 Plain, 12x12 - Log L/40

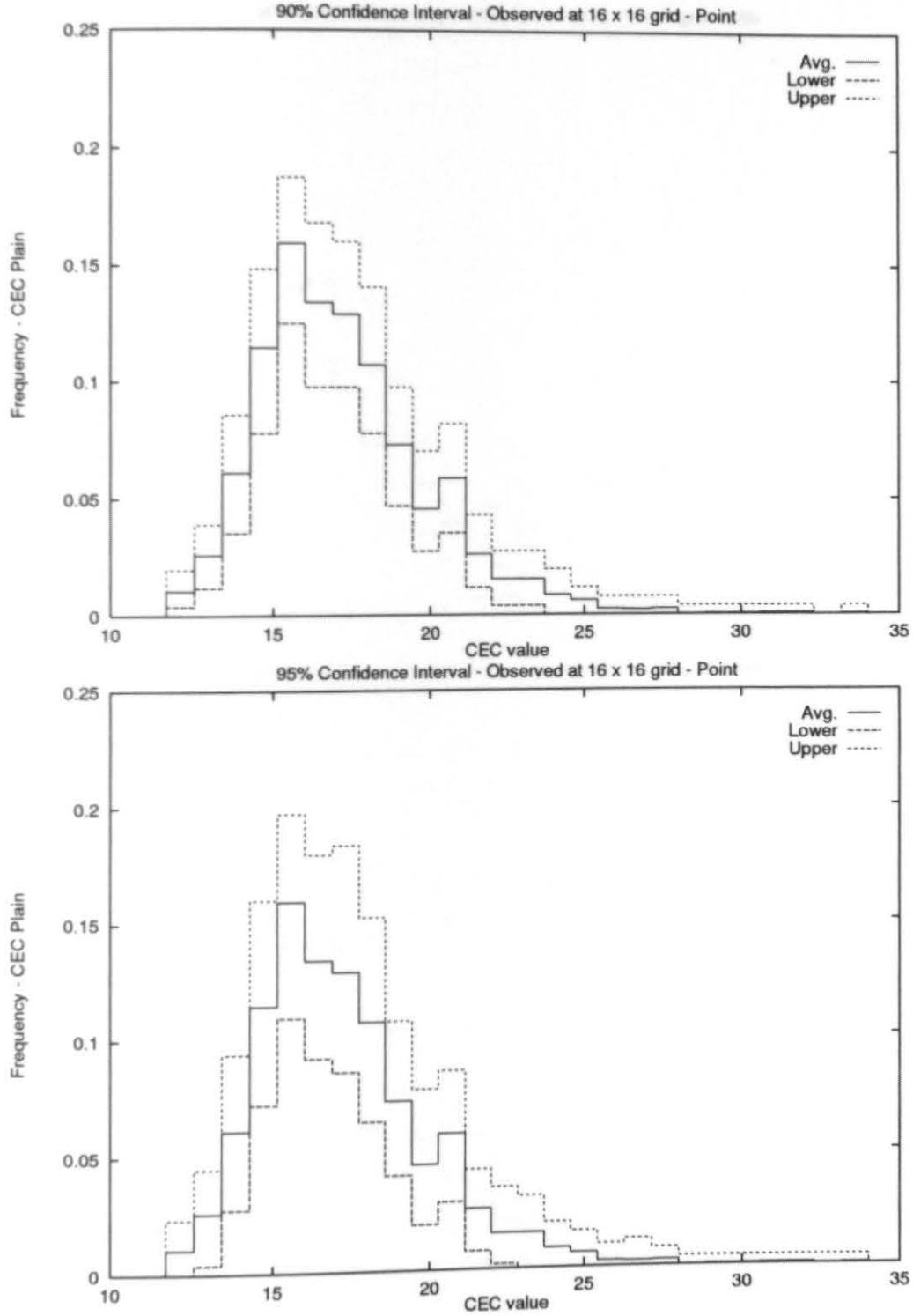


Figure B.42: Histogram Experiment 1 Plain, 16x16 - Obs. point

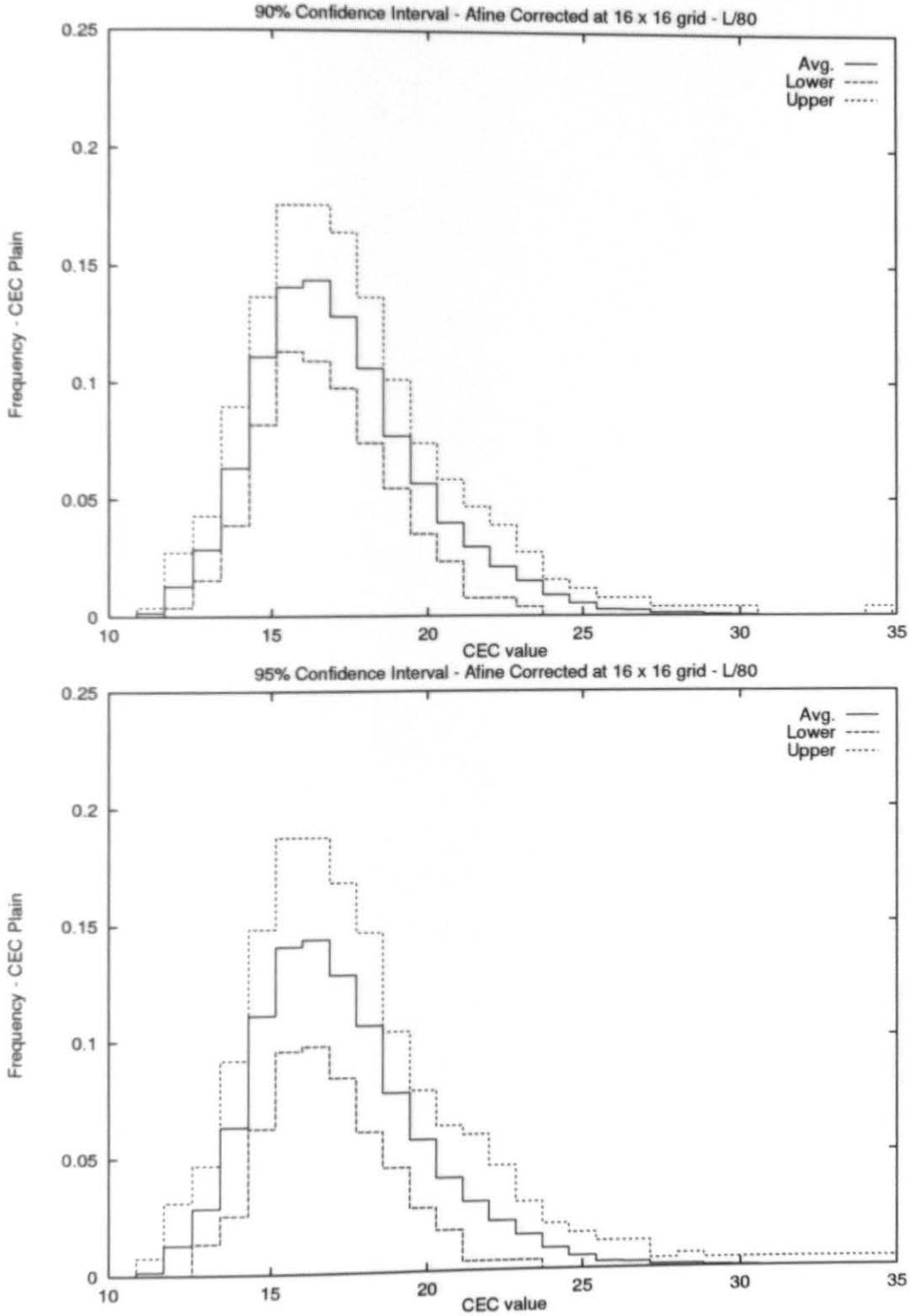


Figure B.43: Histogram Experiment 1 Plain, 16x16 - Afine L/80

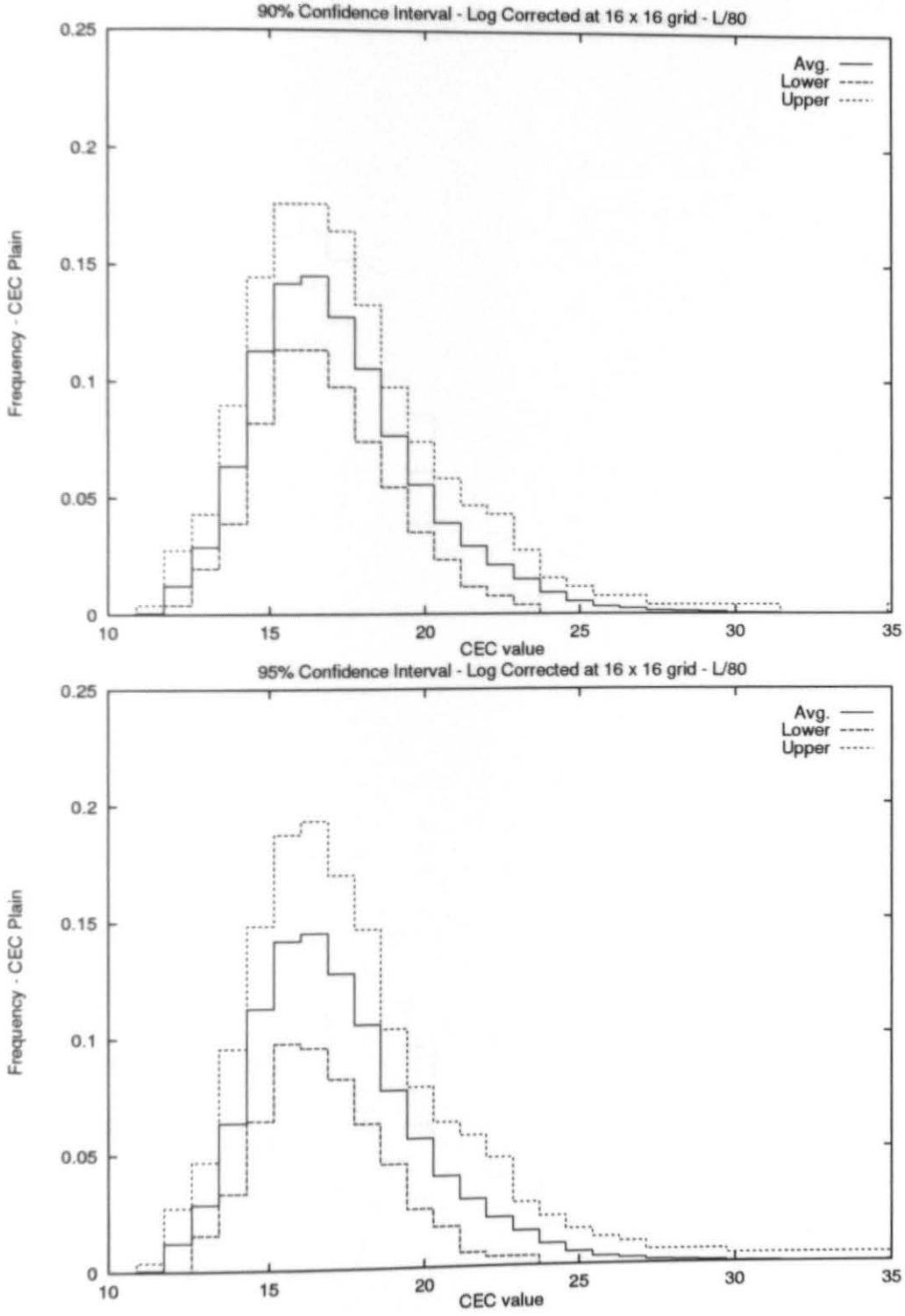


Figure B.44: Histogram Experiment 1 Plain, 16x16 - Log L/80

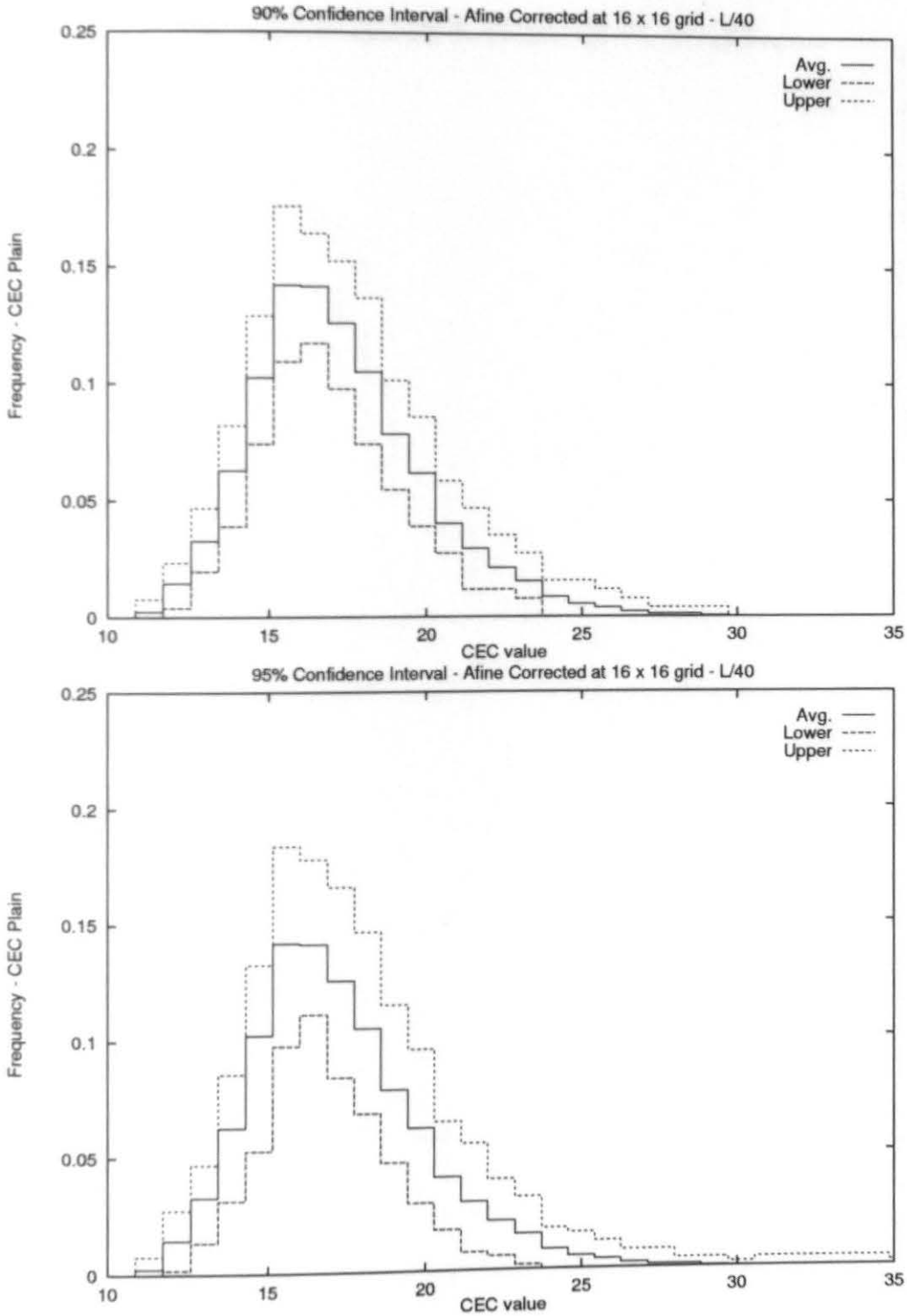


Figure B.45: Histogram Experiment 1 Plain, 16x16 - Afine L/40

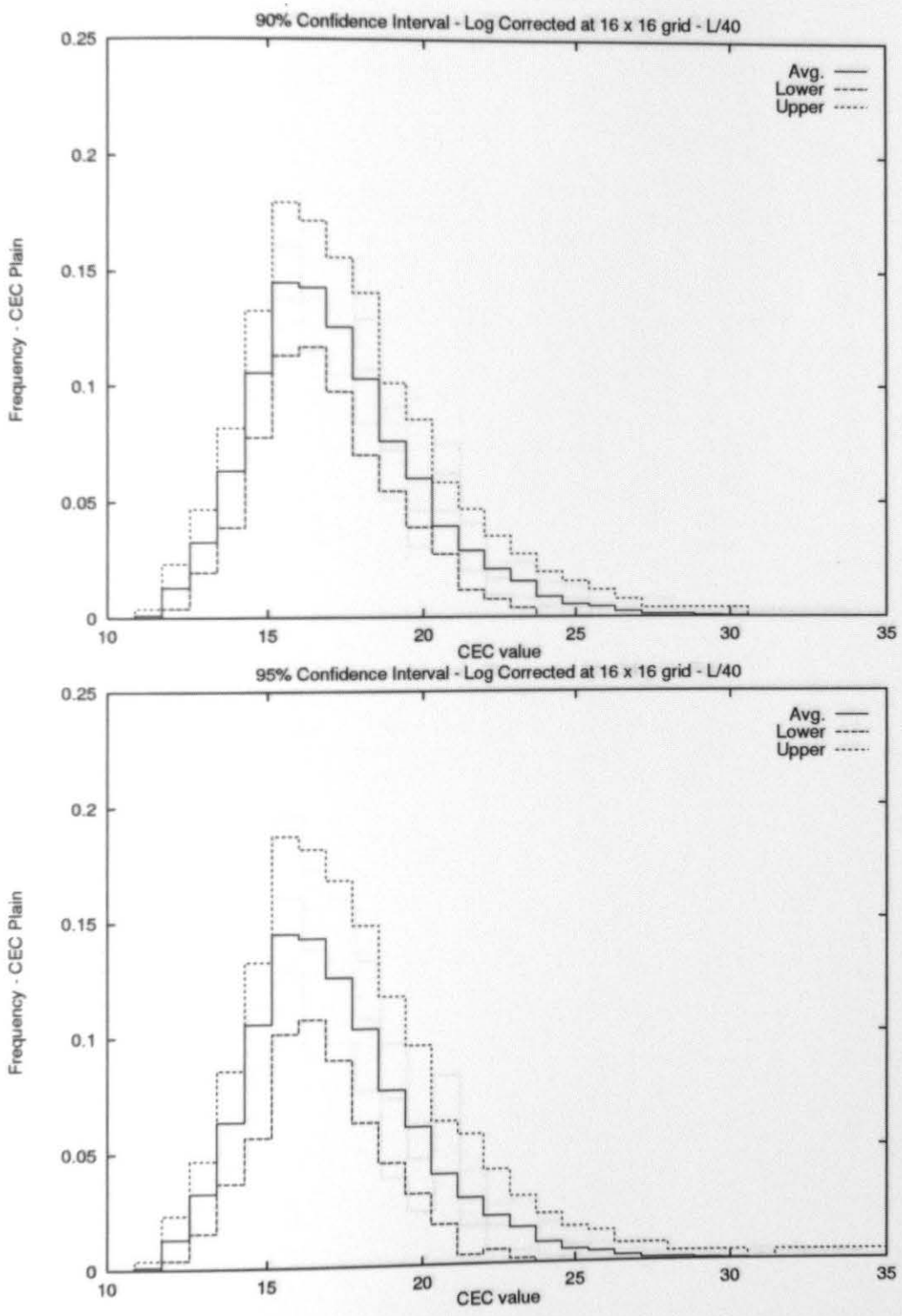


Figure B.46: Histogram Experiment 1 Plain, 16x16 - Log L/40

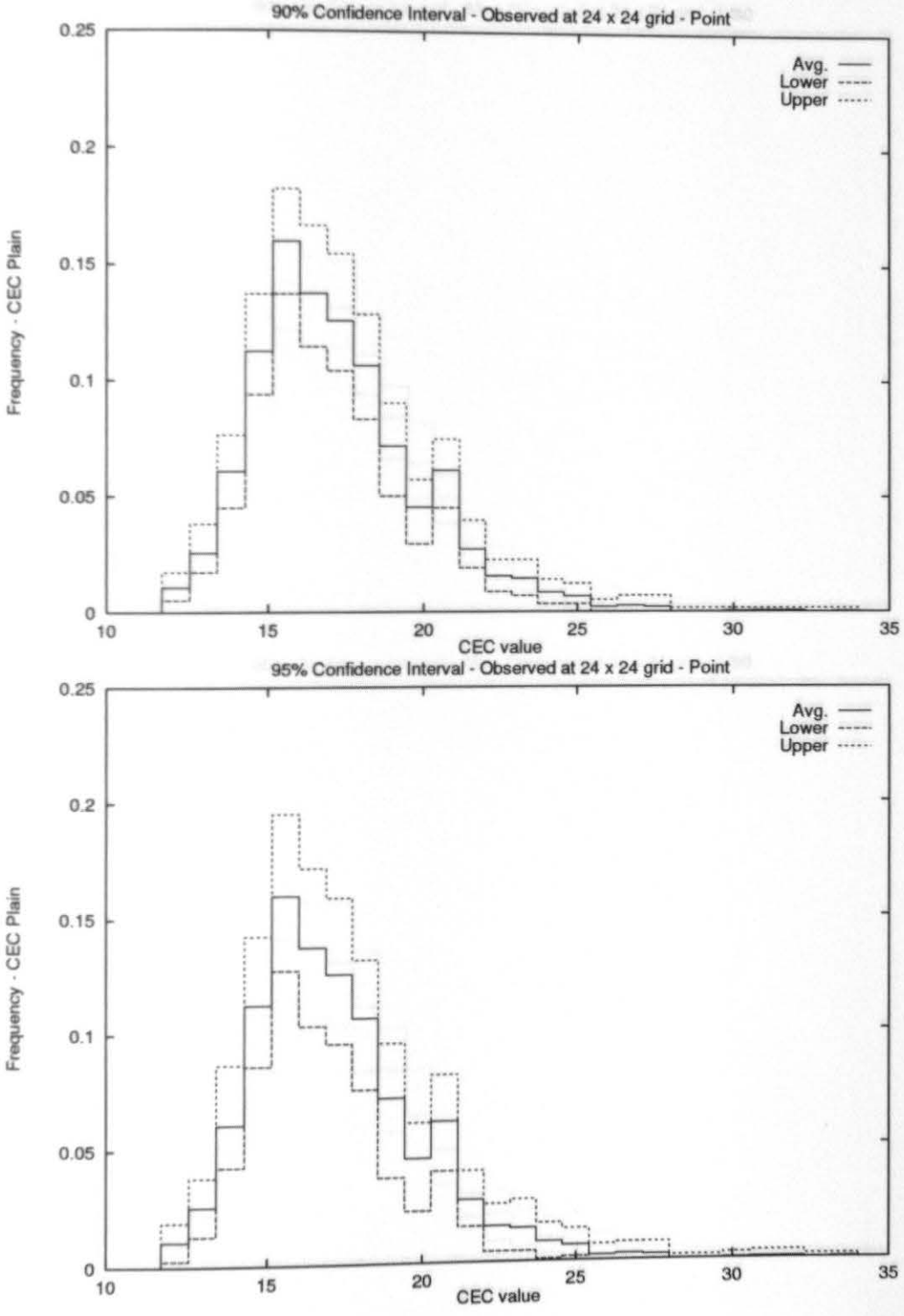


Figure B.47: Histogram Experiment 1 Plain, 24x24 - Obs. point

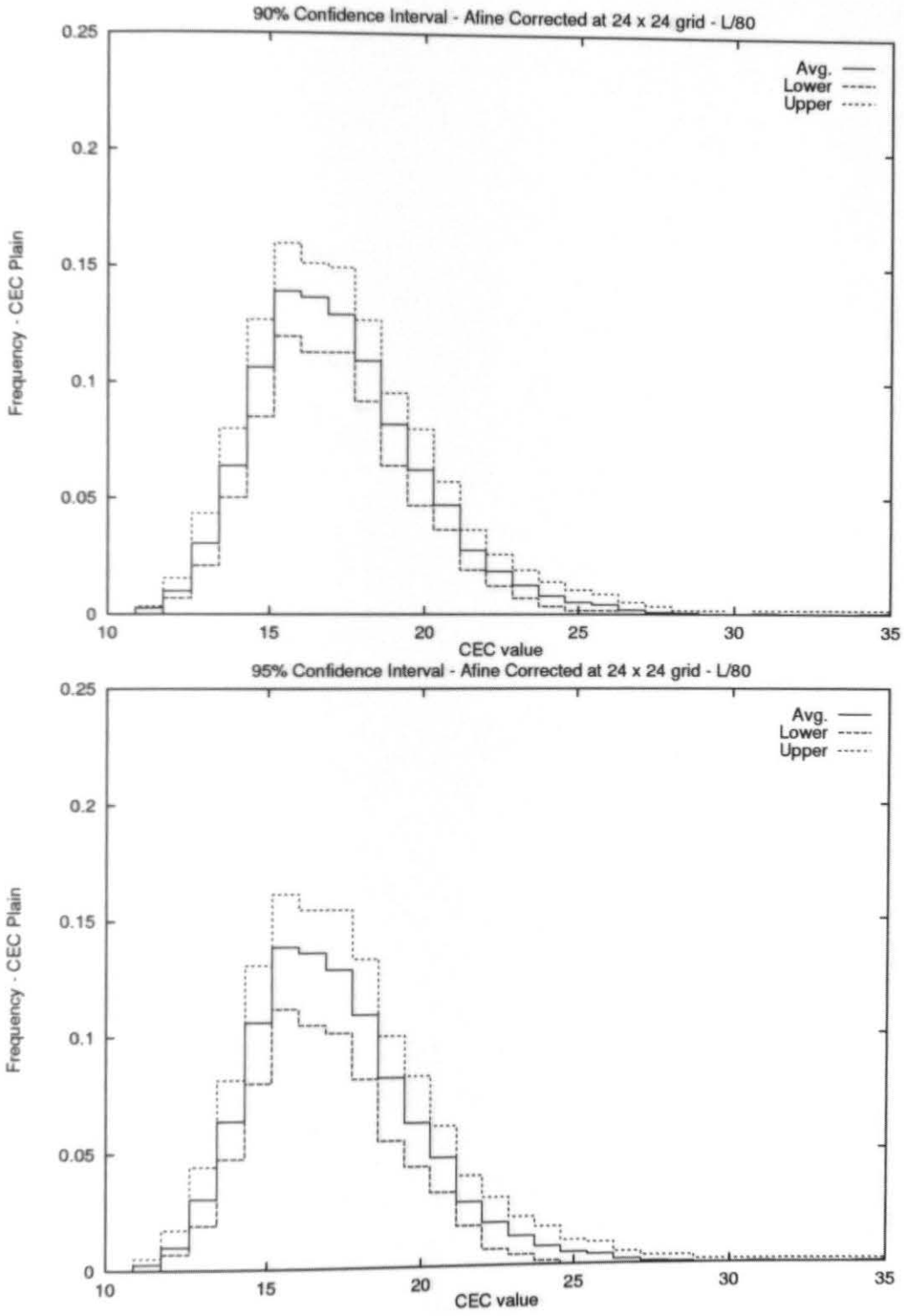


Figure B.48: Histogram Experiment 1 Plain, 24x24 - Afine L/80

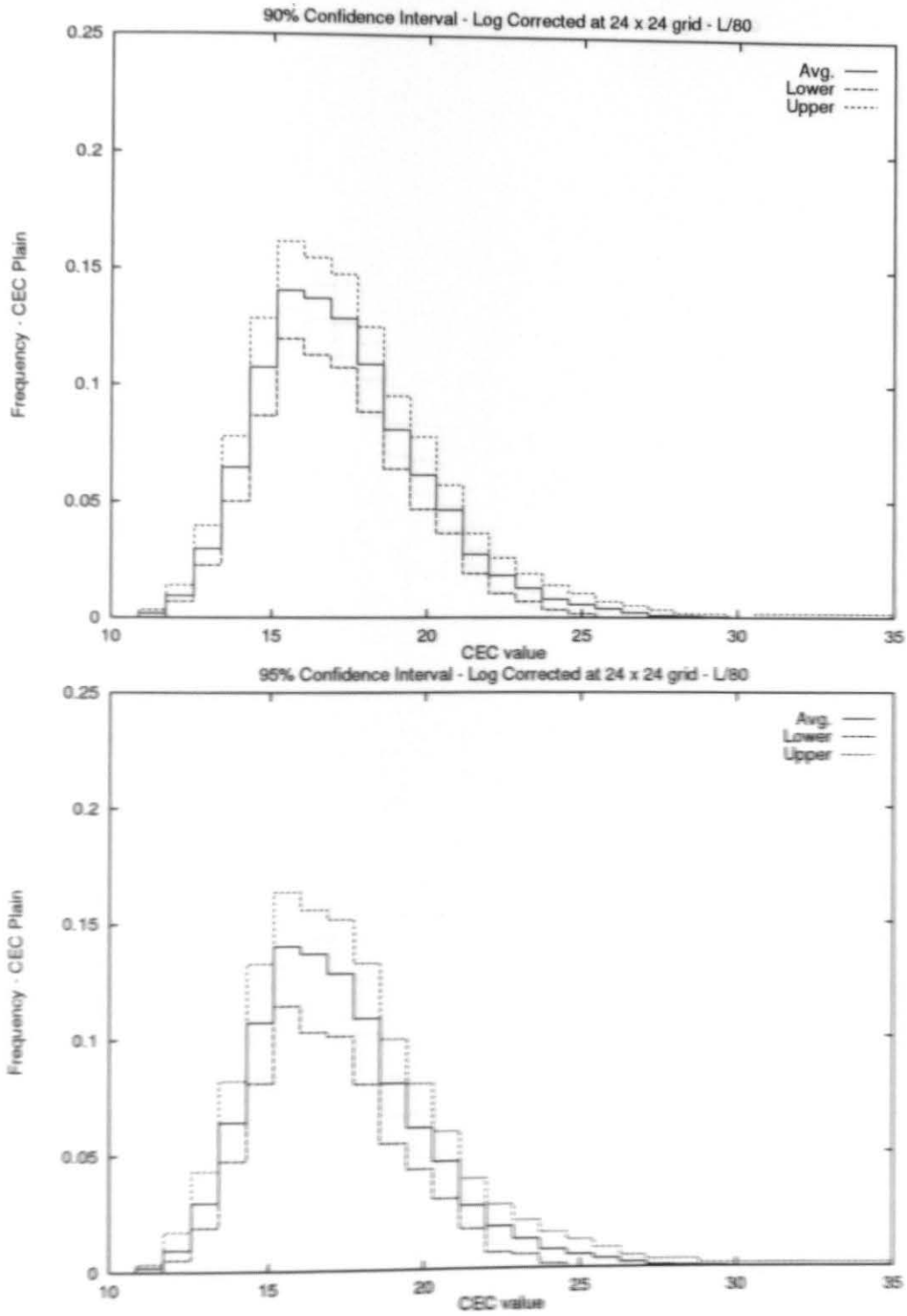


Figure B.49: Histogram Experiment 1 Plain, 24x24 - Log L/80

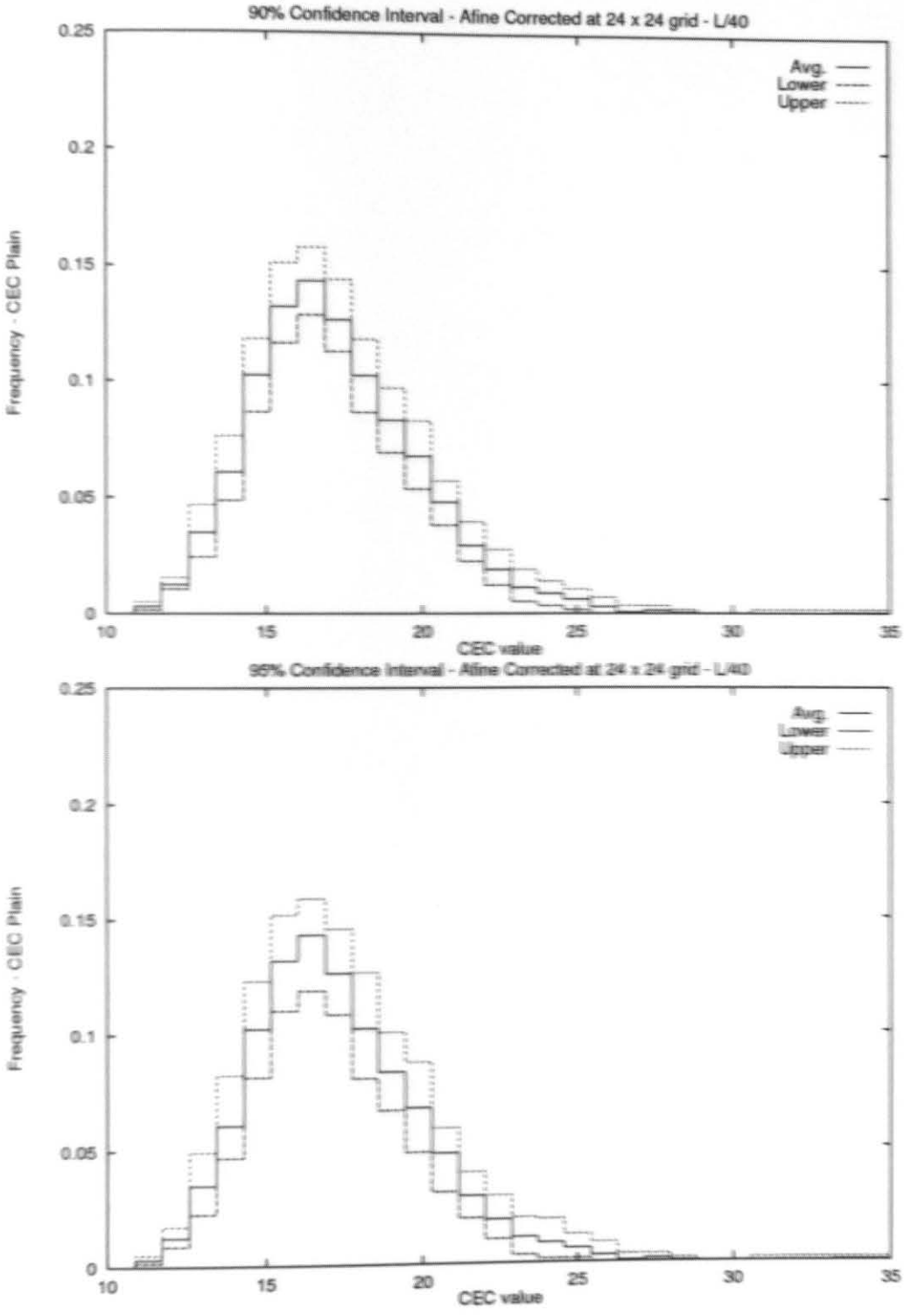


Figure B.50: Histogram Experiment 1 Plain, 24x24 - Afine L/40

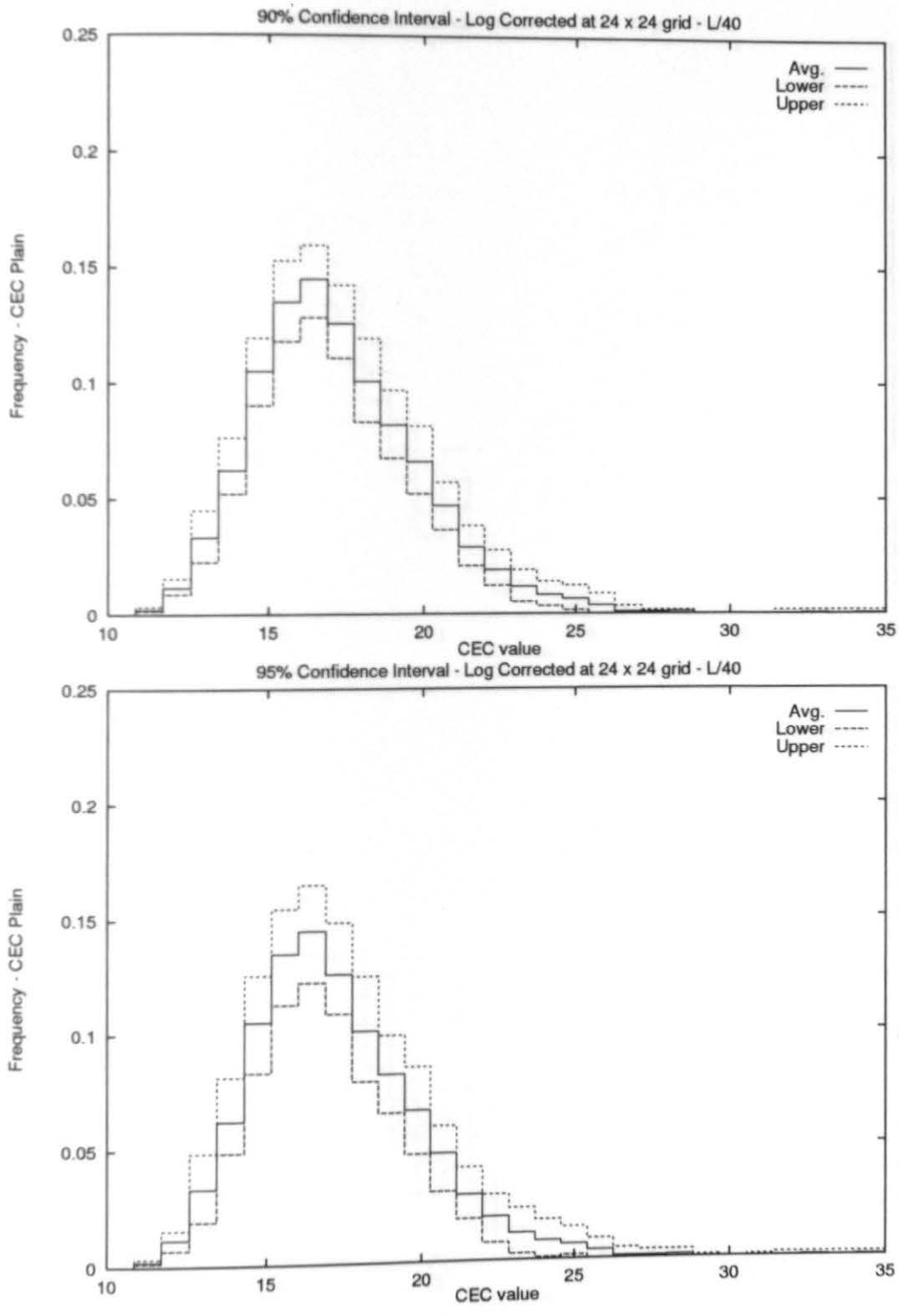


Figure B.51: Histogram Experiment 1 Plain, 24x24 - Log L/40

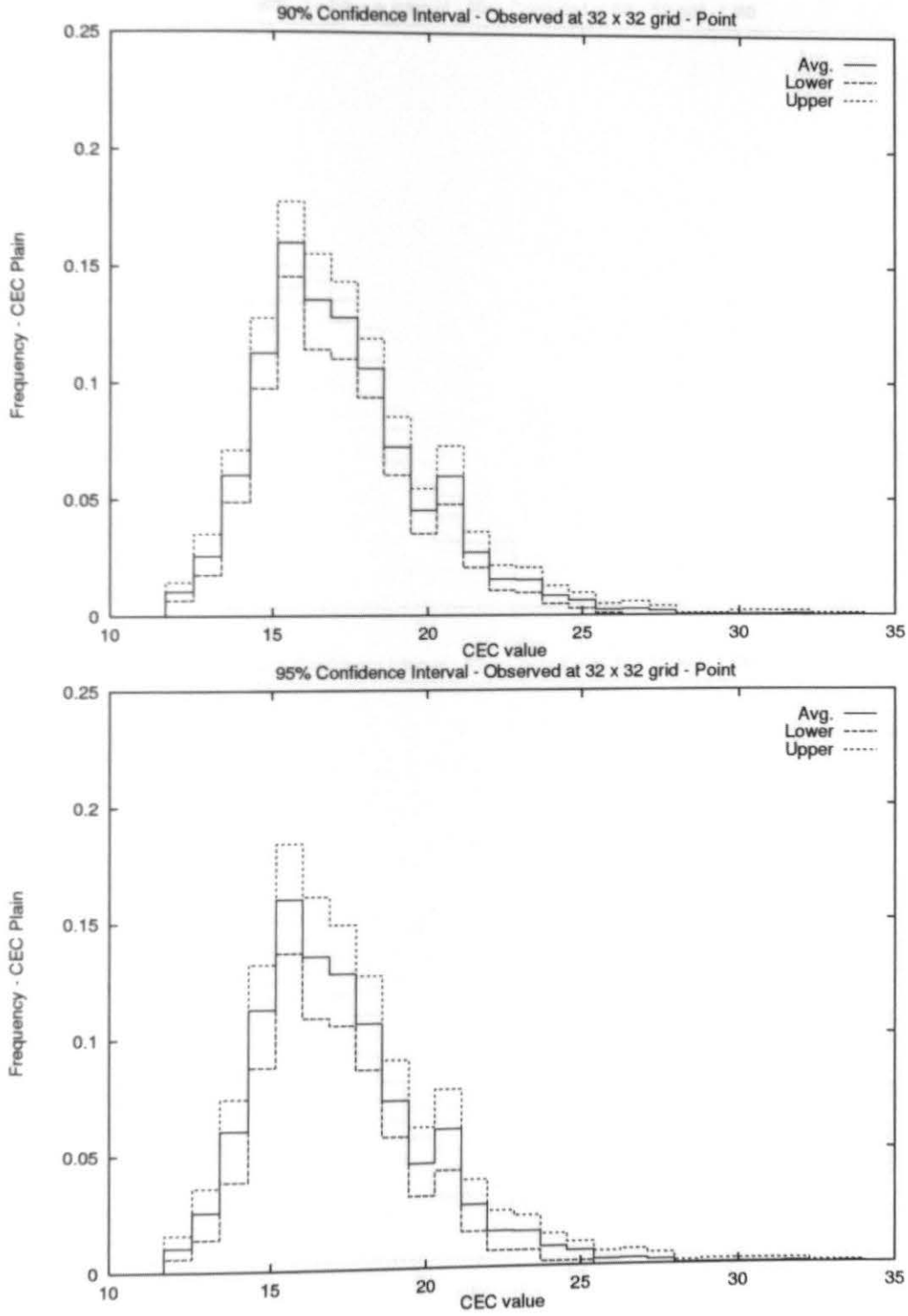


Figure B.52: Histogram Experiment 1 Plain, 32x32 - Obs. point

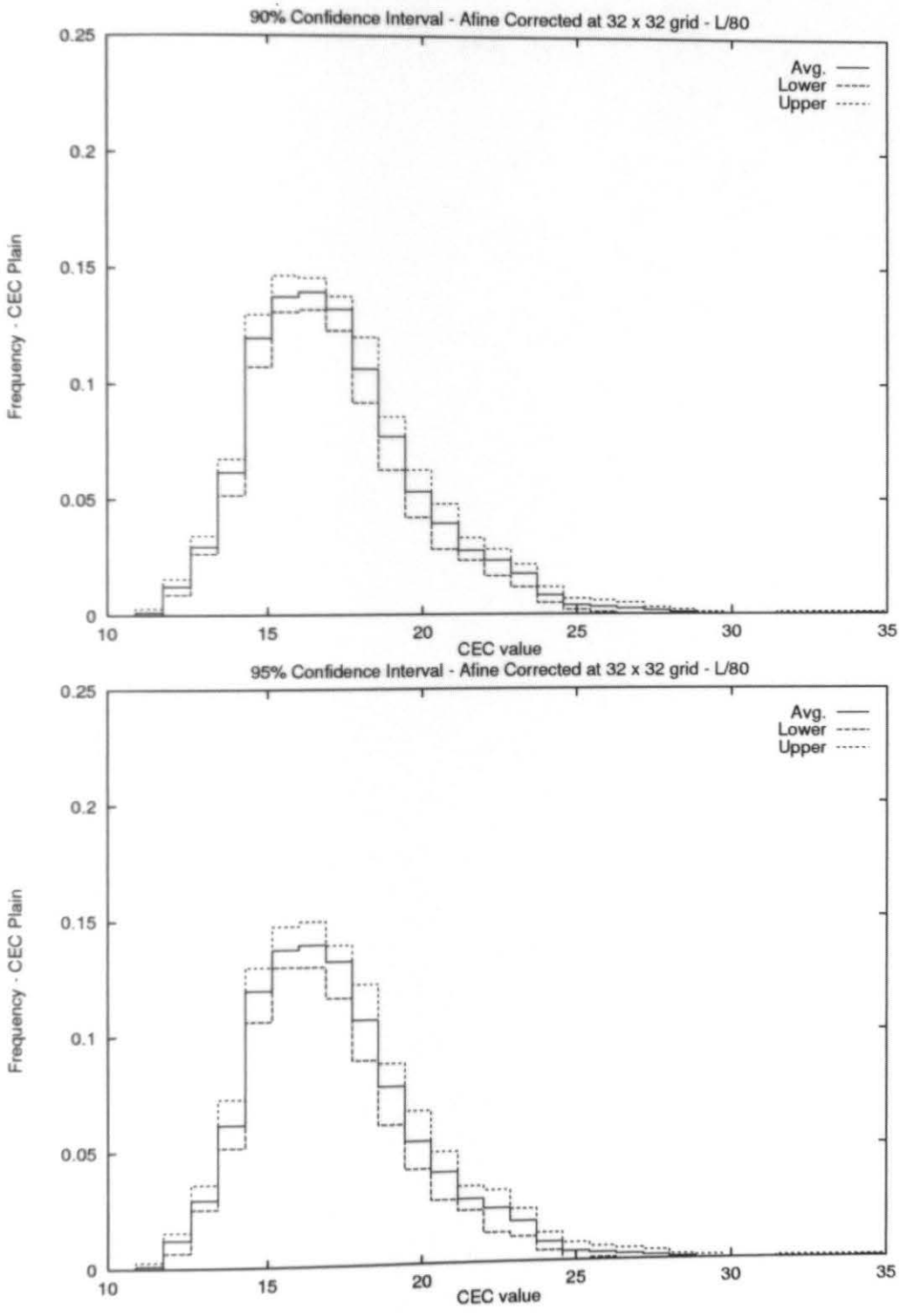


Figure B.53: Histogram Experiment 1 Plain, 32x32 - Afine L/80

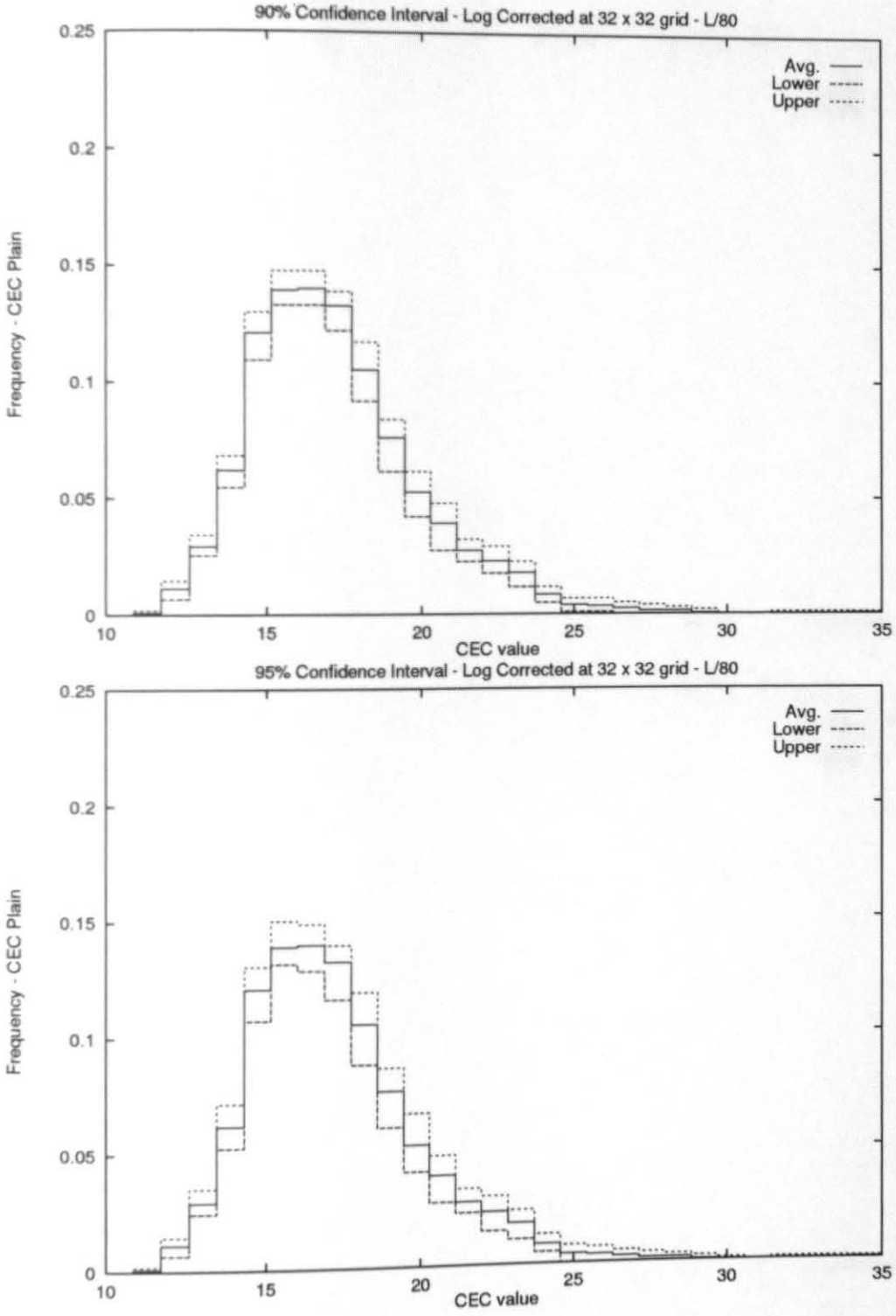


Figure B.54: Histogram Experiment 1 Plain, 32x32 - Log L/80

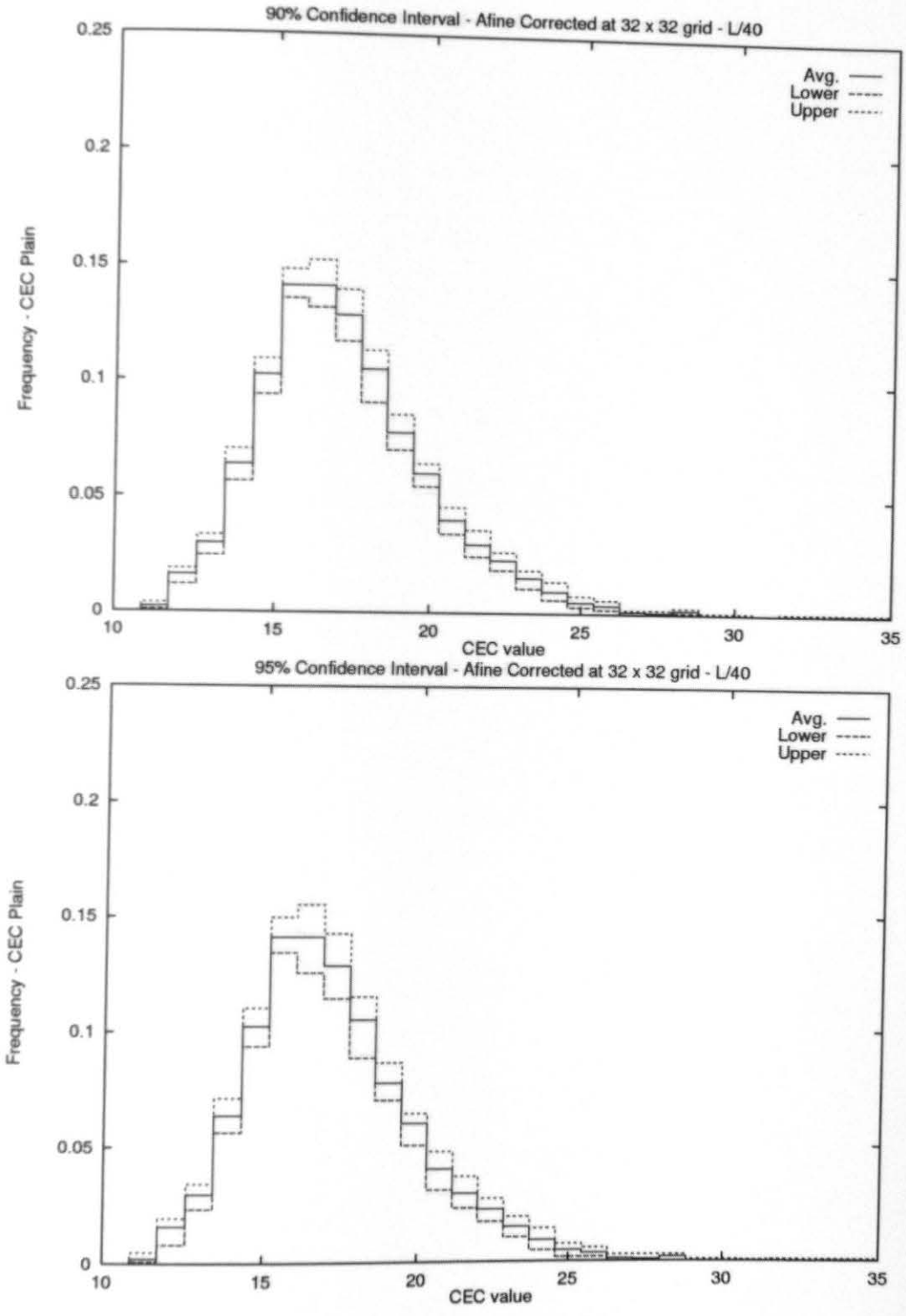


Figure B.55: Histogram Experiment 1 Plain, 32x32 - Afine L/40

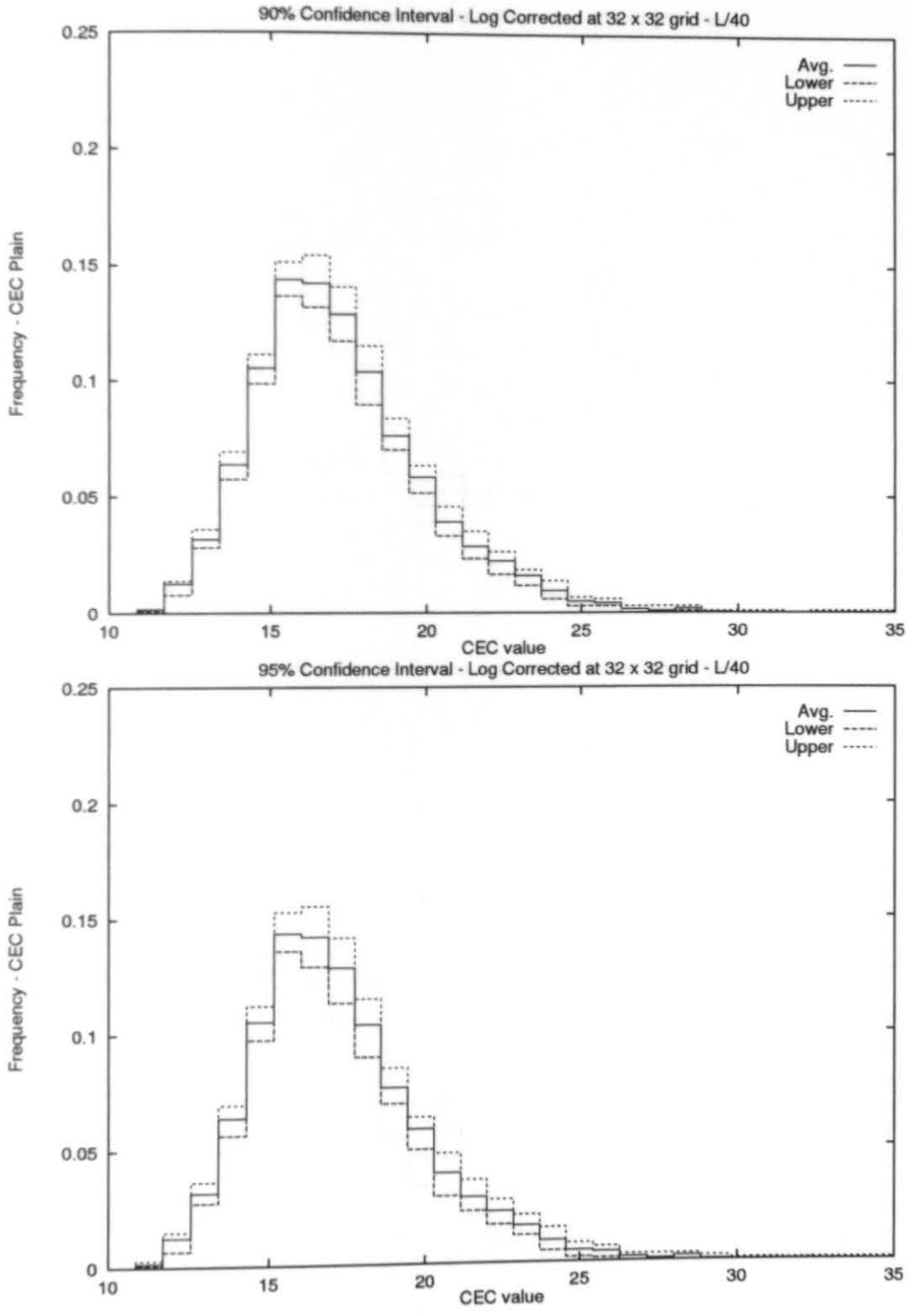


Figure B.56: Histogram Experiment 1 Plain, 32x32 - Log L/40

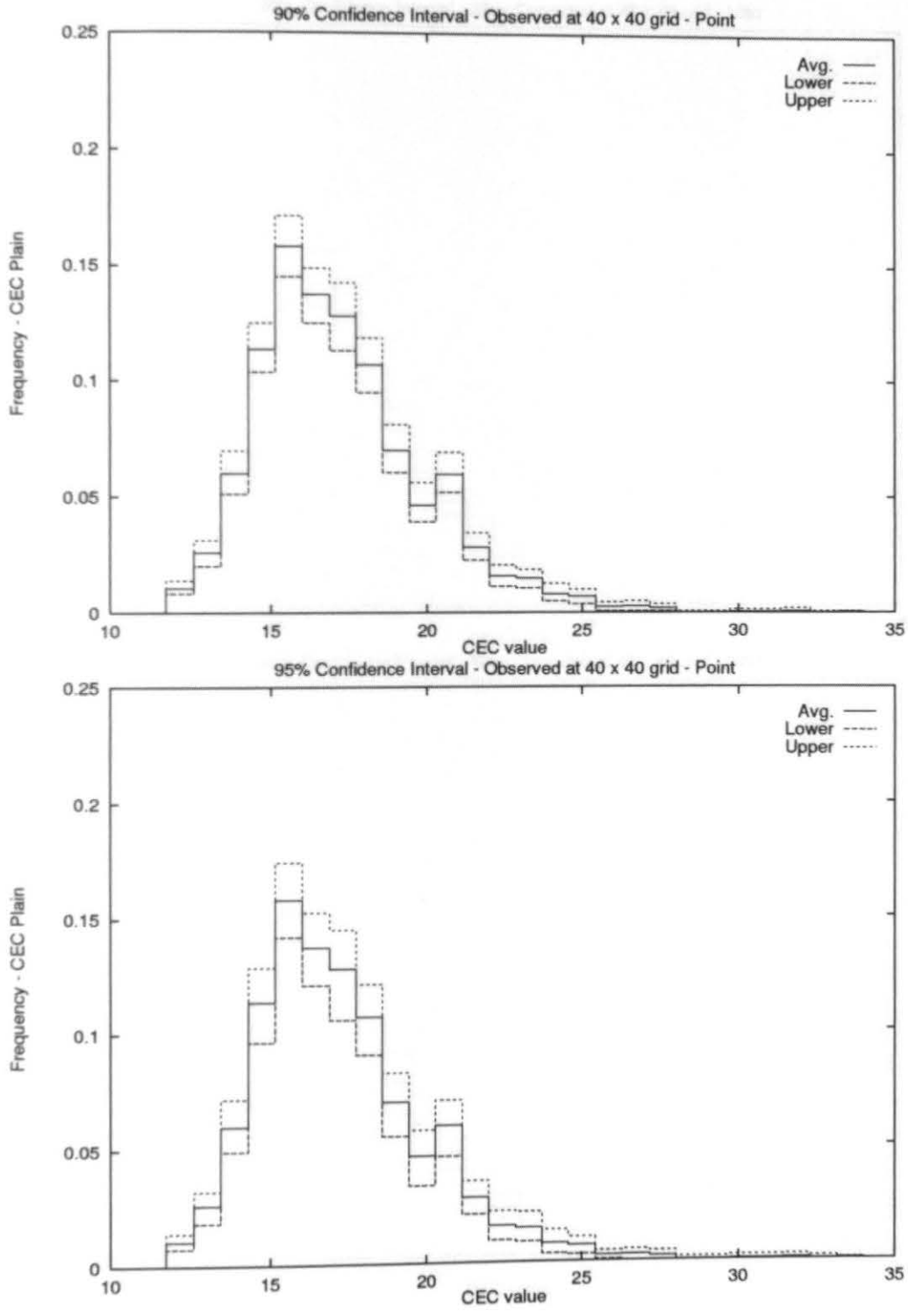


Figure B.57: Histogram Experiment 1 Plain, 40x40 - Obs. point

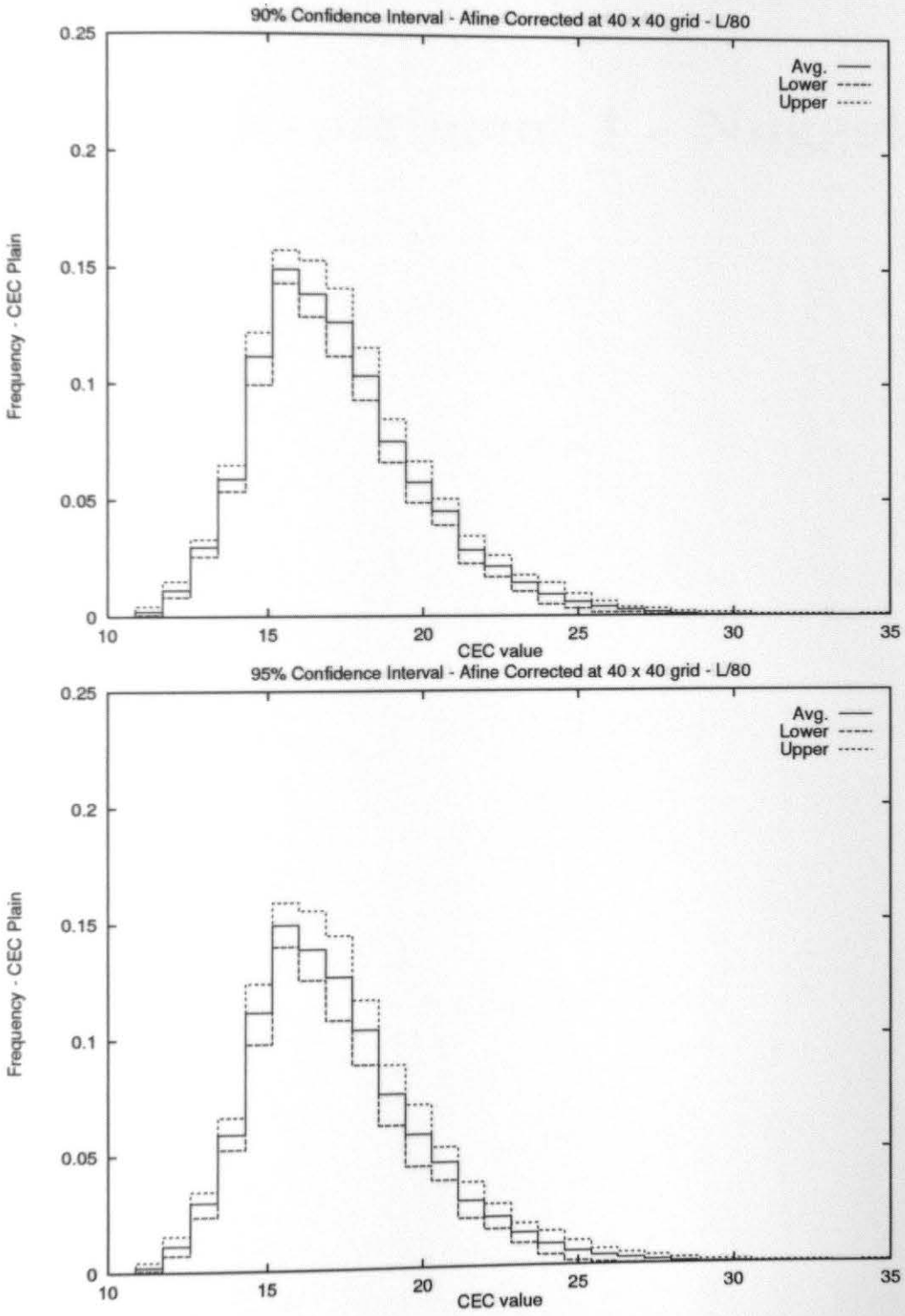


Figure B.58: Histogram Experiment 1 Plain, 40x40 - Afine L/80

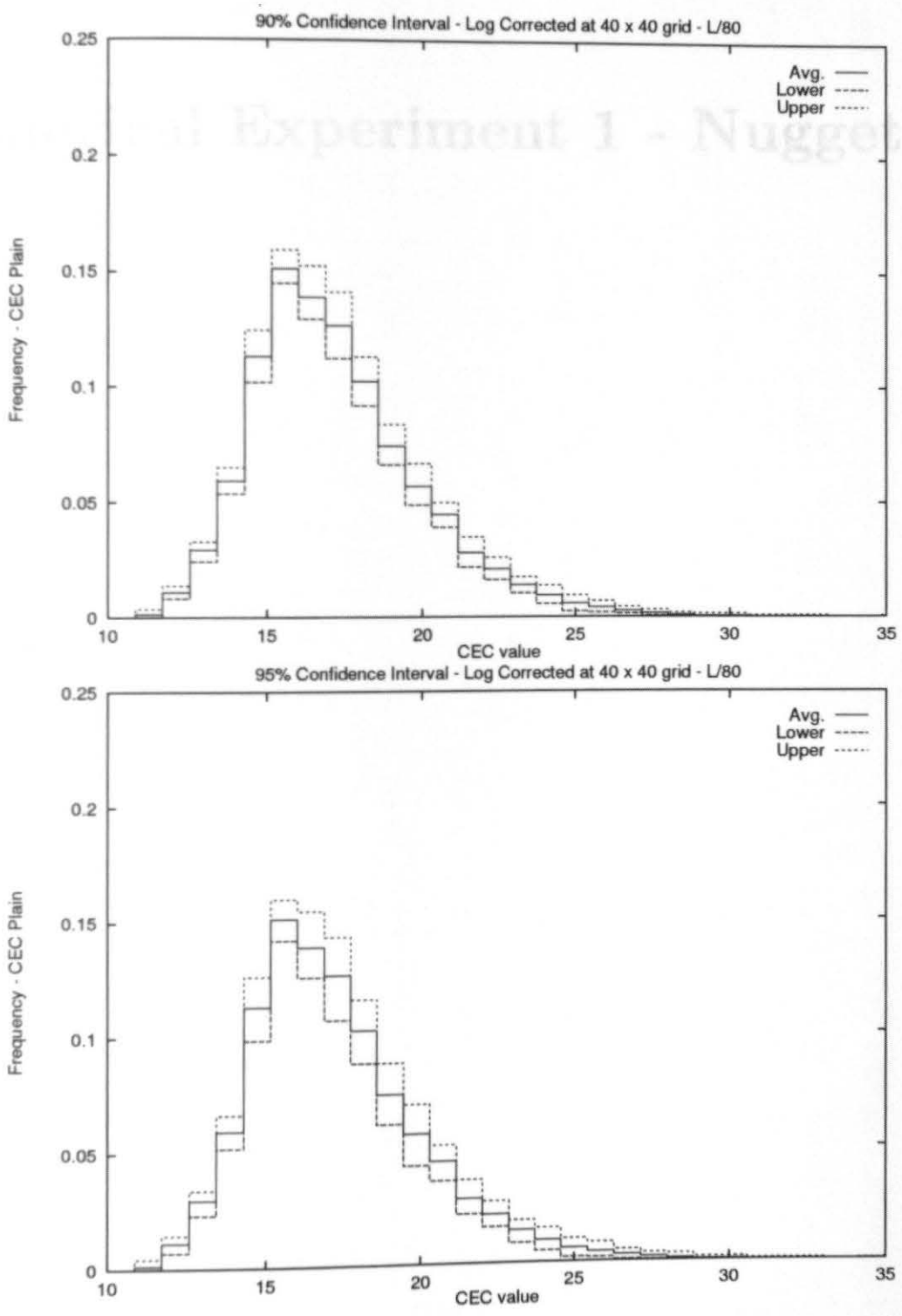


Figure B.59: Histogram Experiment 1 Plain, 40x40 - Log L/80

Numerical Experiment 1 - Nugget

Parameter	Mean	Bias	%Bias	S.D.	%S.D.	C.V.	Low C.I.	High C.I.	S.D. C.I.	Low C.I.	High C.I.
Point:											
Sill	5.831	0.142	2.438	1.171	20.082	0.201	5.811	5.851	0.026	3.775	8.878
Range	73.870	11.686	15.819	18.027	24.403	0.244	-331.903	479.643	369.323	46.353	107.601
L/80:											
Sill	5.599	0.090	1.615	0.864	15.430	0.154	5.580	5.617	0.038	4.420	7.597
Range	44.915	17.269	38.448	36.334	80.895	0.809	-18.662	108.492	143.991	4.520	101.619
L/40:											
Sill	5.472	0.217	3.959	0.812	14.831	0.148	5.429	5.516	0.077	4.200	7.199
Range	44.251	17.933	40.526	34.929	78.934	0.789	-1.466	89.968	96.847	4.583	95.525
Mixed:											
Sill	5.555	0.134	2.409	0.827	14.895	0.149	5.528	5.583	0.055	4.382	7.442
Range	44.153	18.031	40.838	35.751	80.971	0.810	-2.216	90.521	108.240	4.539	101.565
Case A:											
Sill	5.893	0.204	3.462	1.054	17.880	0.179	5.881	5.905	0.024	4.276	8.510
Range	35.597	26.587	74.688	36.537	102.641	1.026	-11.245	82.440	181.344	4.772	113.930
Case B:											
Sill	5.797	0.108	1.871	1.022	17.622	0.176	5.788	5.807	0.025	4.333	8.453
Range	28.829	33.355	115.701	32.473	112.642	1.126	18.088	39.569	49.824	4.239	101.450
Case C:											
Sill	5.810	0.121	2.081	1.008	17.355	0.174	5.803	5.817	0.018	4.367	8.431
Range	28.736	33.448	116.397	32.217	112.112	1.121	21.895	35.577	30.654	4.364	97.057

Table C.1: Experiment 1, Nugget, 12 x 12, sill = 5.689, range = 3*62.184. Confidence Intervals (C.I.) are at the 95% level

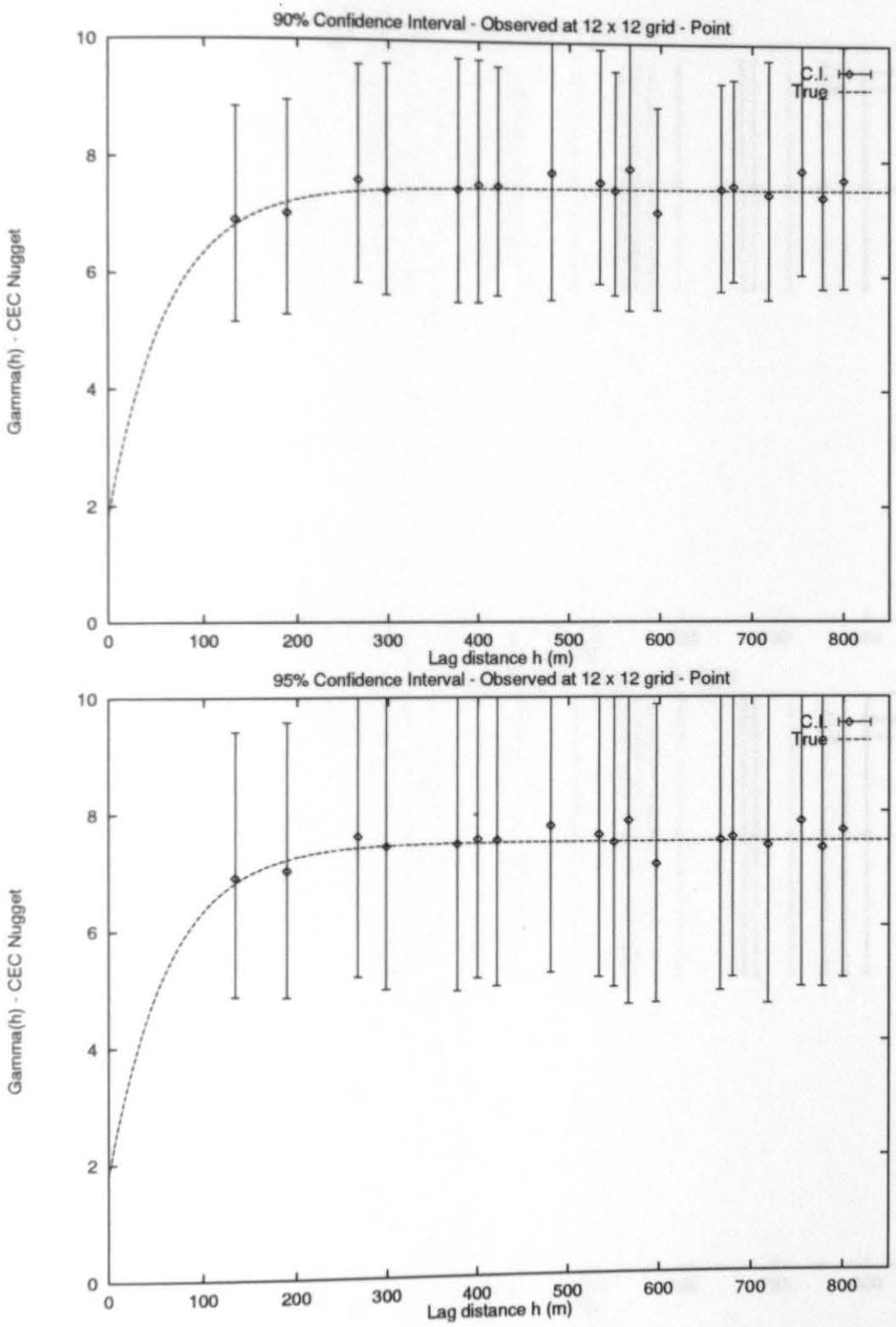


Figure C.1: Numerical Experiment 1 Nugget, 12x12 - Obs. Point

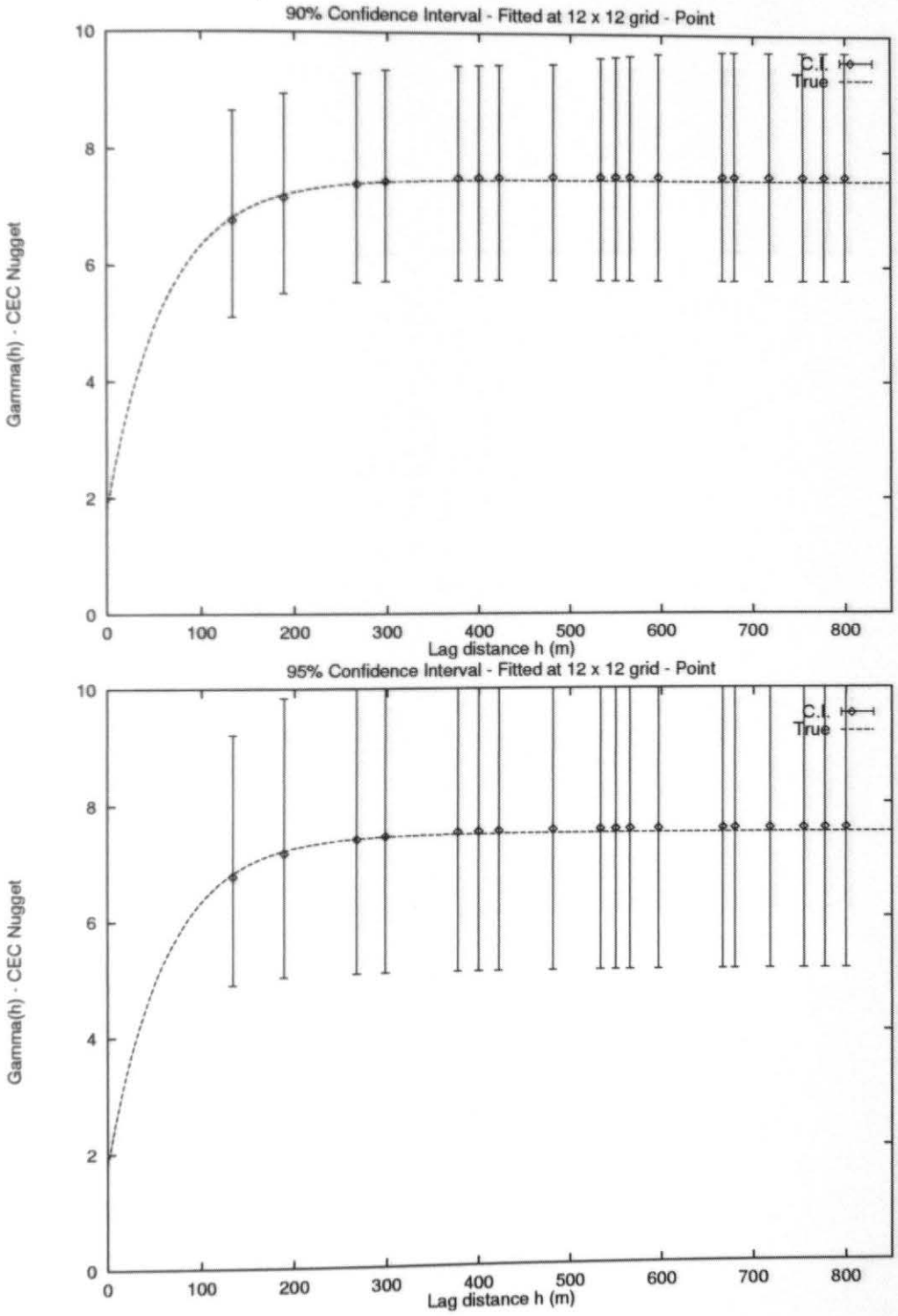


Figure C.2: Numerical Experiment 1 Nugget, 12x12 - Fit. Point

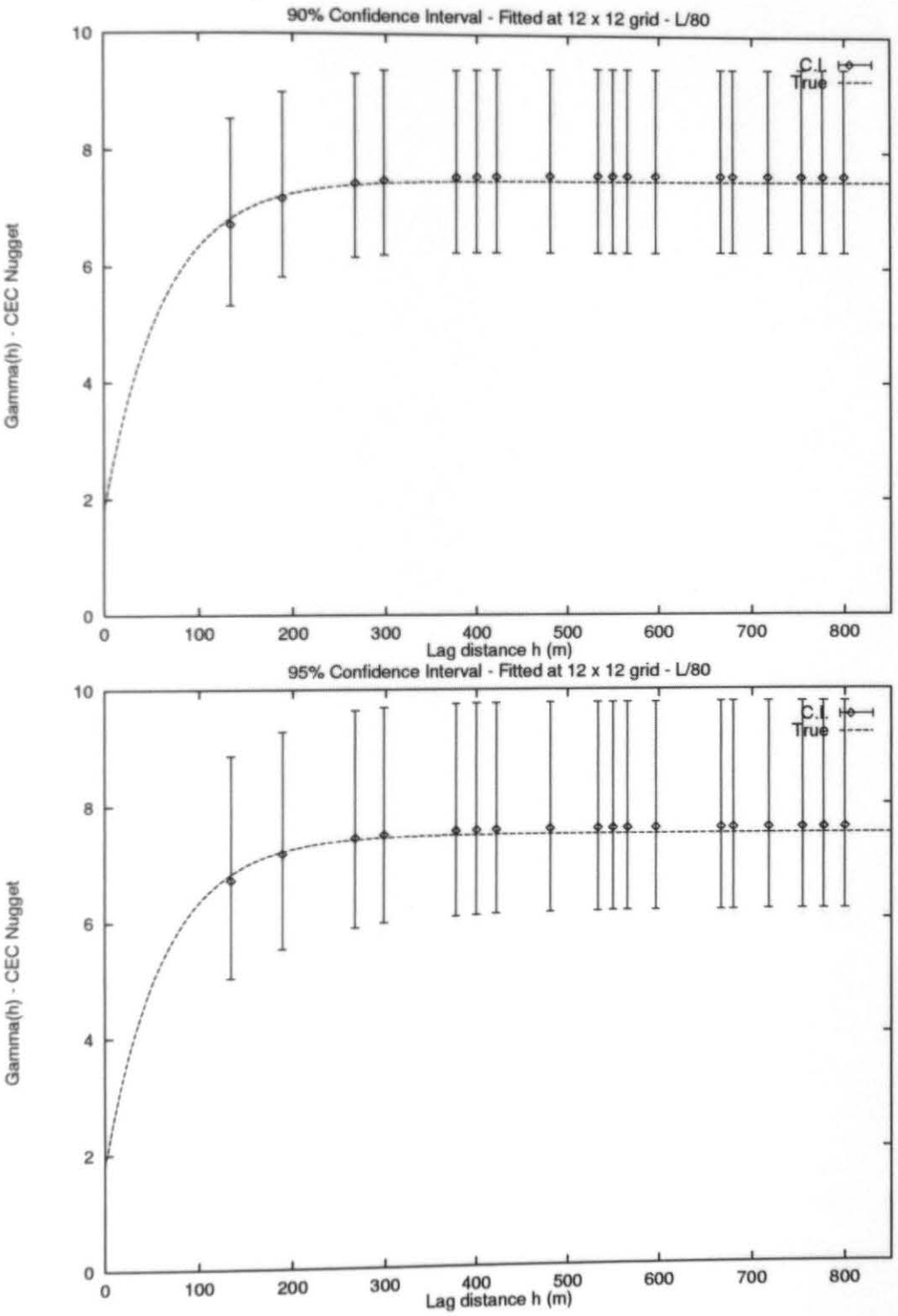


Figure C.3: Numerical Experiment 1 Nugget, 12x12 - L/80

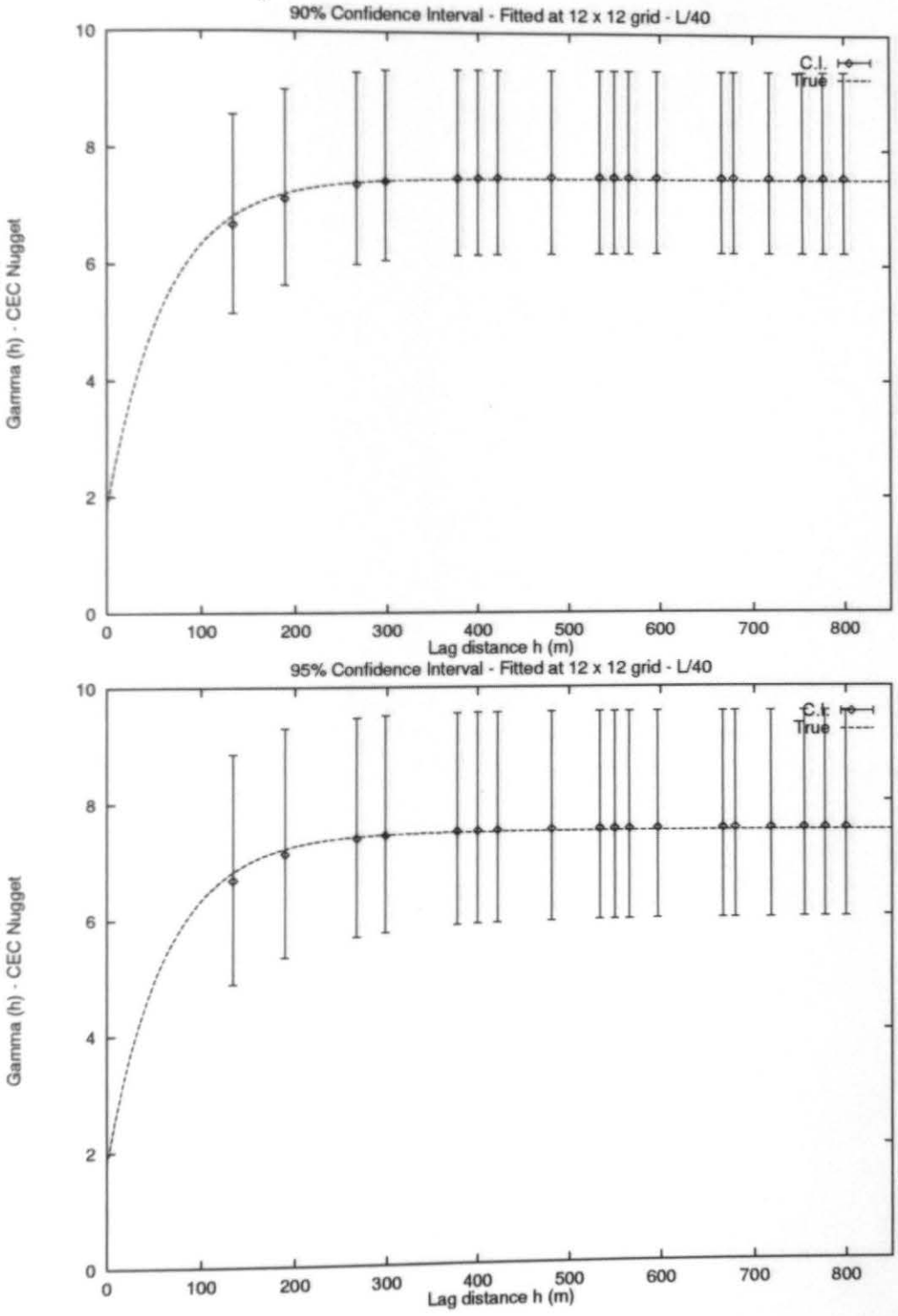


Figure C.4: Numerical Experiment 1 Nugget, 12x12 - L/40

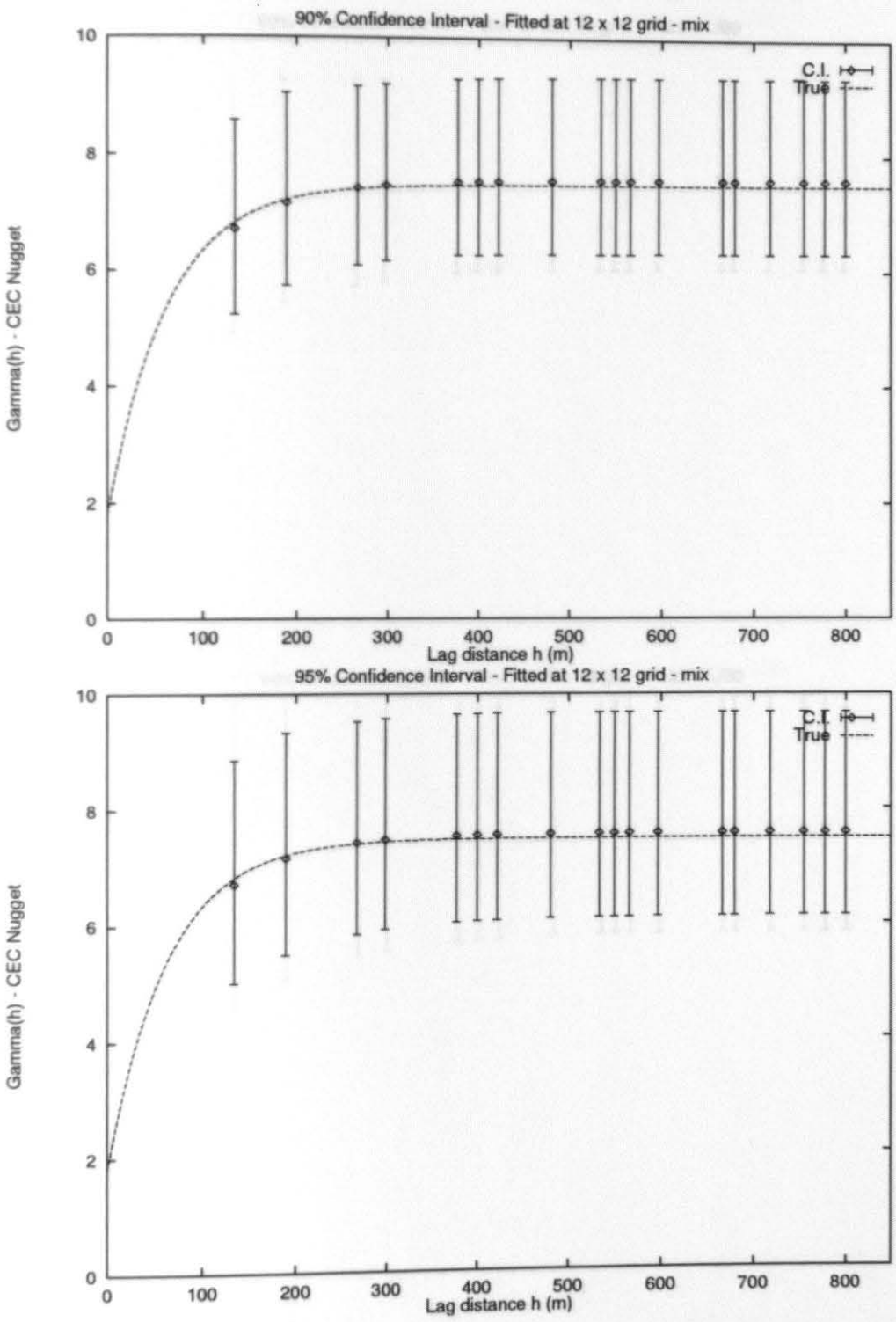


Figure C.5: Numerical Experiment 1 Nugget, 12x12 - mix

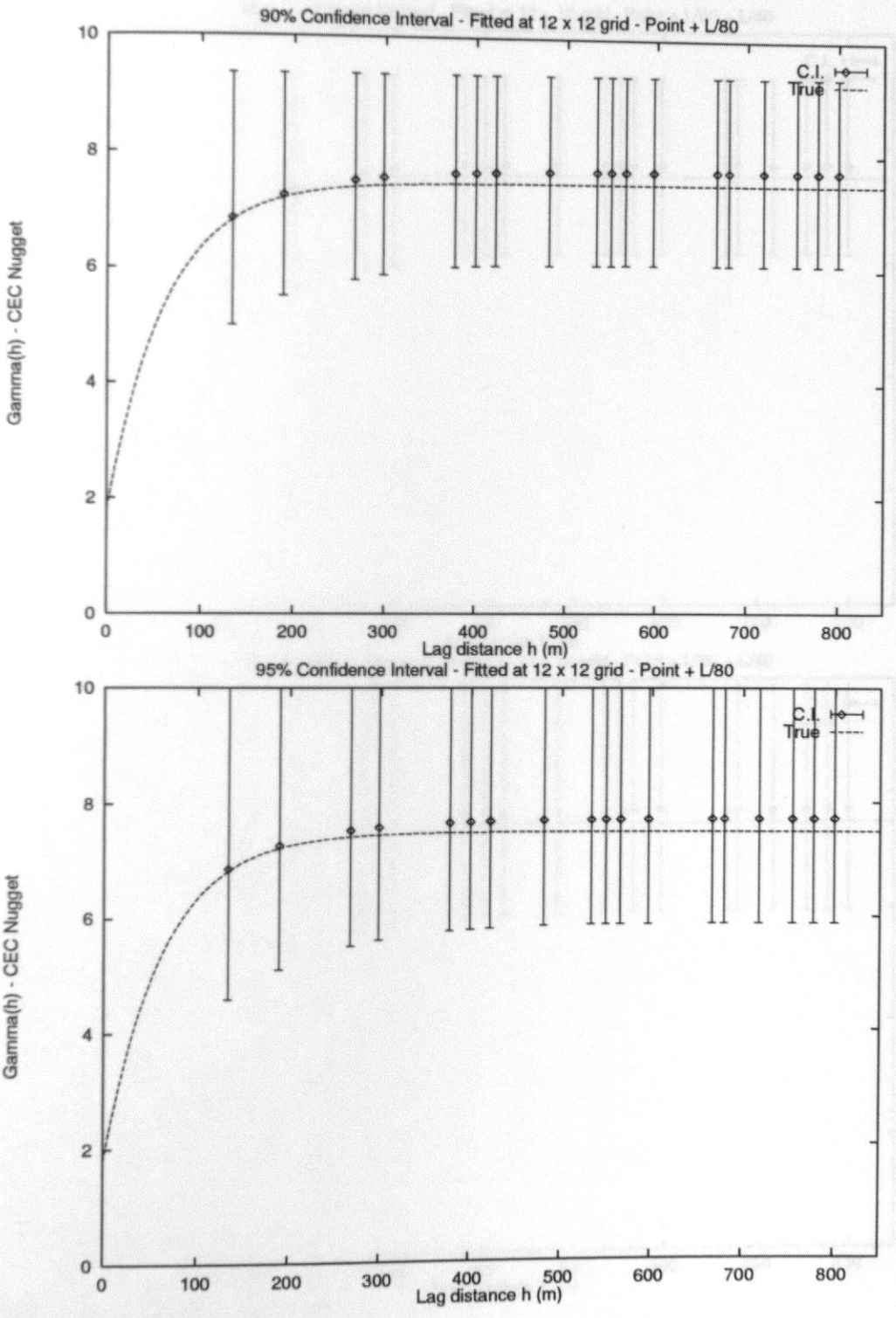


Figure C.6: Numerical Experiment 1 Nugget, 12x12 - point + L/80

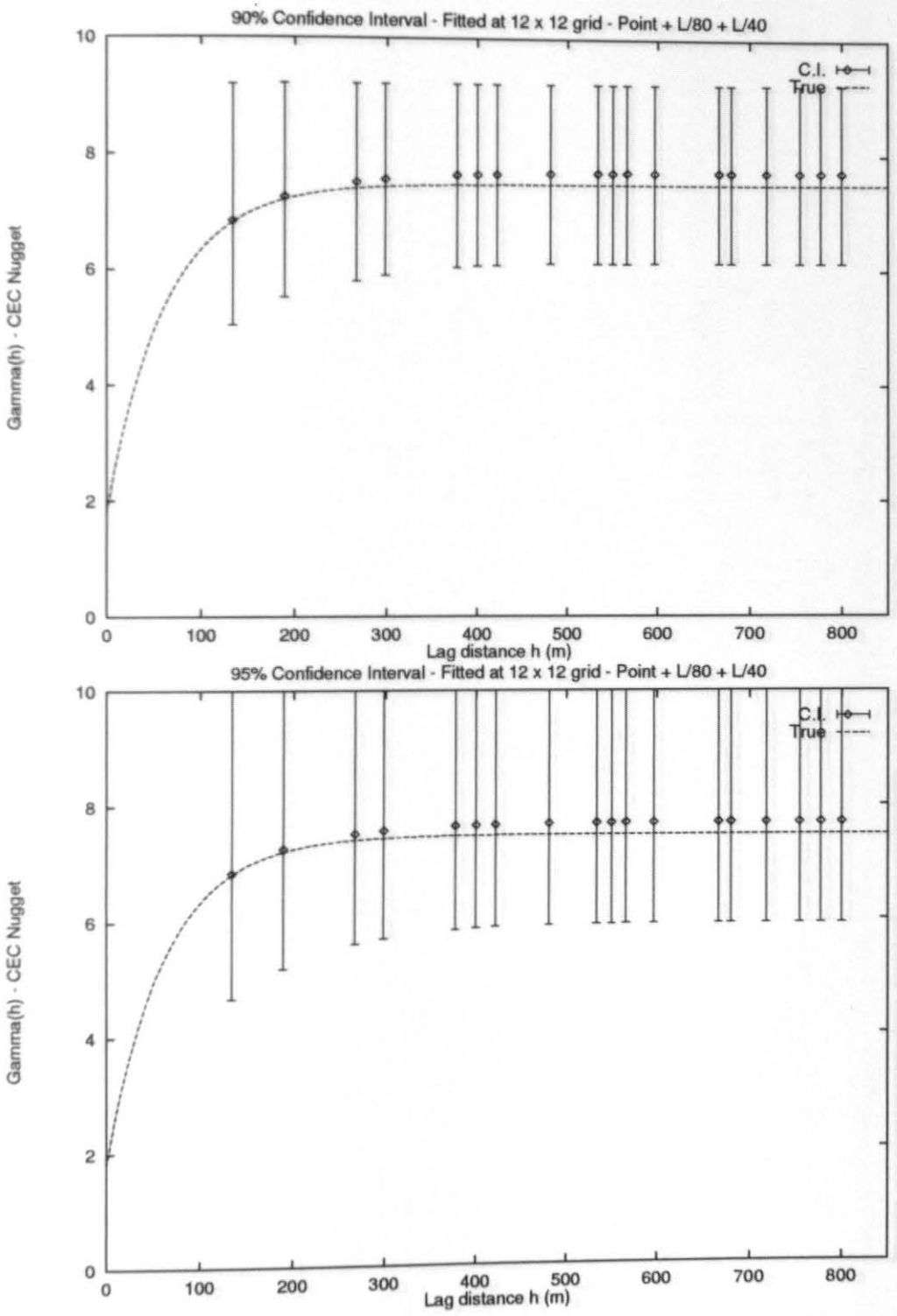


Figure C.7: Numerical Experiment 1 Nugget, 12x12 - point + L/80 + L/40

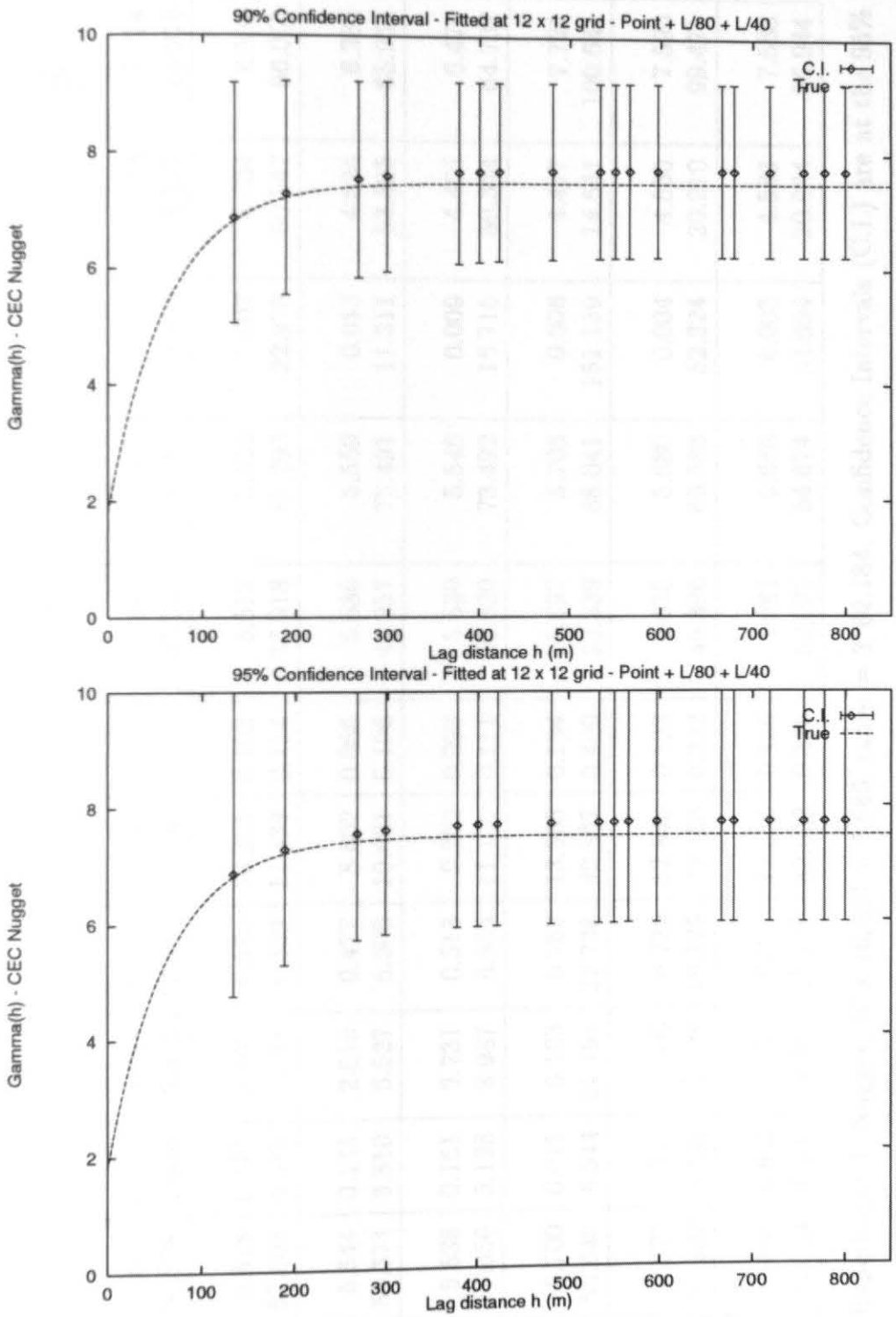


Figure C.8: Numerical Experiment 1 Nugget, 12x12 - point + L/80 + L/40 + mix

Parameter	Mean	Bias	%Bias	S.D.	%S.D.	C.V.	Low C.I.	High C.I.	S.D. C.I.	Low C.I.	High C.I.
Point:											
Sill	5.650	0.039	0.684	0.872	15.431	0.154	5.646	5.655	0.006	3.943	7.774
Range	60.578	1.606	2.652	12.955	21.386	0.214	-27.233	148.389	124.412	33.332	85.915
L/80:											
Sill	5.518	0.171	3.108	0.570	10.325	0.103	5.512	5.524	0.007	4.304	6.556
Range	57.605	4.579	7.948	6.589	11.439	0.114	33.918	81.293	22.978	50.547	66.052
L/40:											
Sill	5.544	0.145	2.610	0.477	8.602	0.086	5.530	5.559	0.013	4.624	6.384
Range	58.374	3.810	6.527	6.200	10.621	0.106	41.257	75.491	11.311	53.505	65.954
Mixed:											
Sill	5.538	0.151	2.731	0.513	9.262	0.093	5.530	5.546	0.009	4.465	6.475
Range	57.056	5.128	8.987	6.354	11.136	0.111	40.620	73.492	16.715	50.344	64.737
Case A:											
Sill	5.700	0.011	0.193	0.762	13.365	0.134	5.697	5.703	0.006	4.477	7.721
Range	55.640	6.544	11.761	22.738	40.867	0.409	23.239	88.041	151.159	14.641	100.641
Case B:											
Sill	5.677	0.012	0.206	0.729	12.840	0.128	5.675	5.680	0.004	4.530	7.590
Range	57.992	4.192	7.228	19.229	33.158	0.332	46.396	69.588	52.324	20.210	99.471
Case C:											
Sill	5.686	0.003	0.051	0.717	12.616	0.126	5.684	5.688	0.003	4.583	7.588
Range	57.384	4.800	8.365	18.720	32.622	0.326	50.095	64.674	31.694	20.064	96.944

Table C.2: Experiment 1, Nugget, 16 x 16, sill = 5.689, range = 3*62.184. Confidence Intervals (C.I.) are at the 95% level

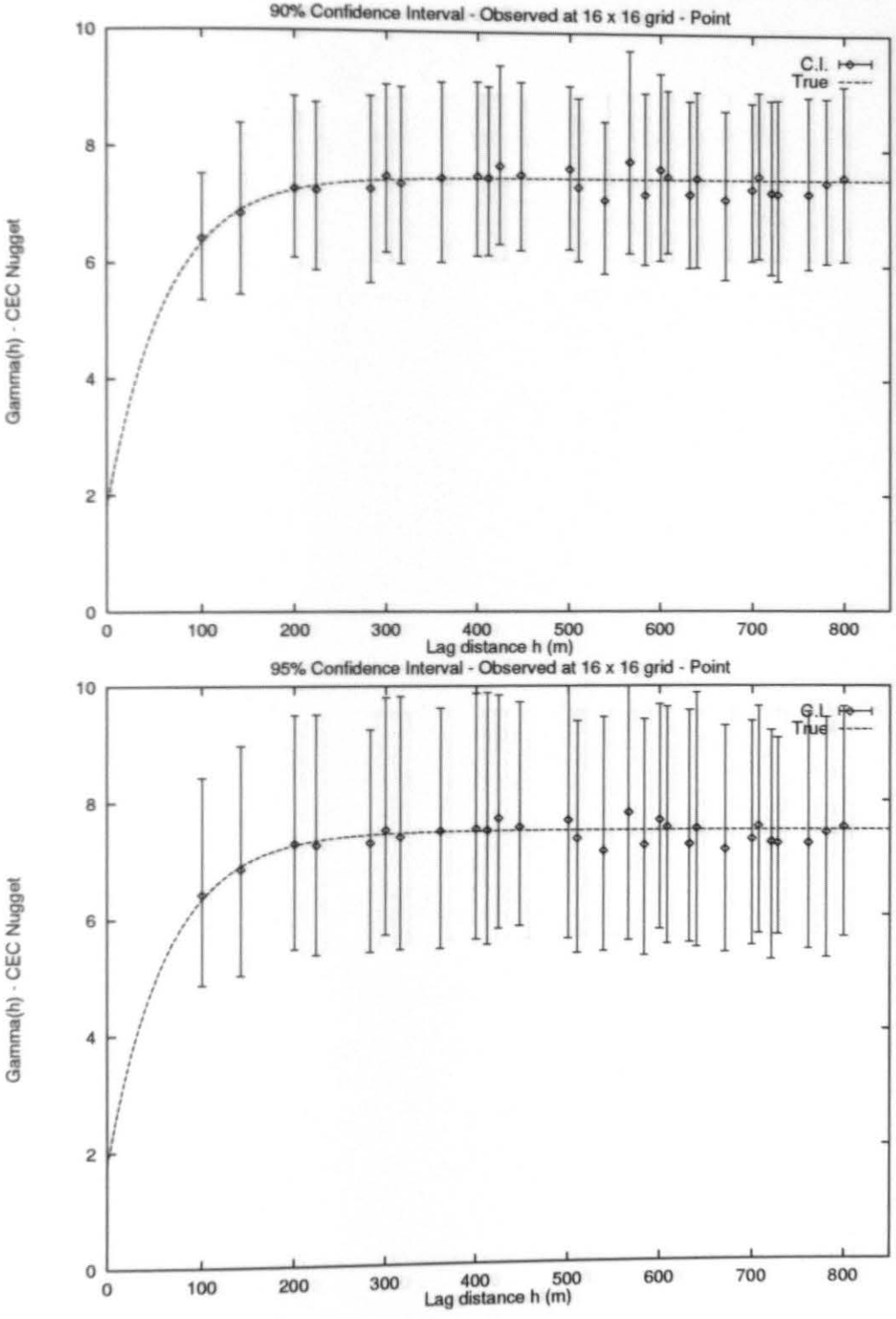


Figure C.9: Numerical Experiment 1 Nugget, 16x16 - Obs. Point

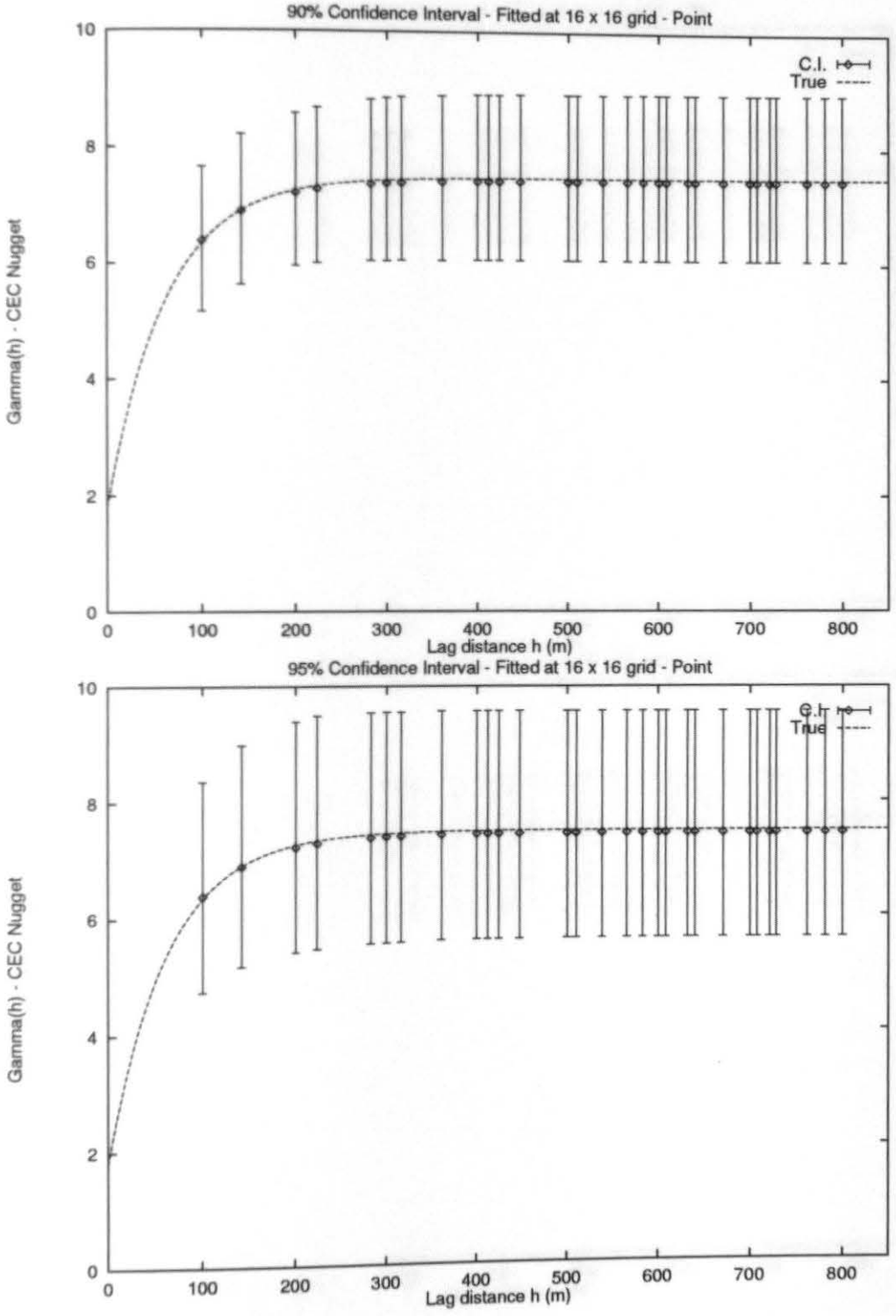


Figure C.10: Numerical Experiment 1 Nugget, 16x16 - Fit. Point

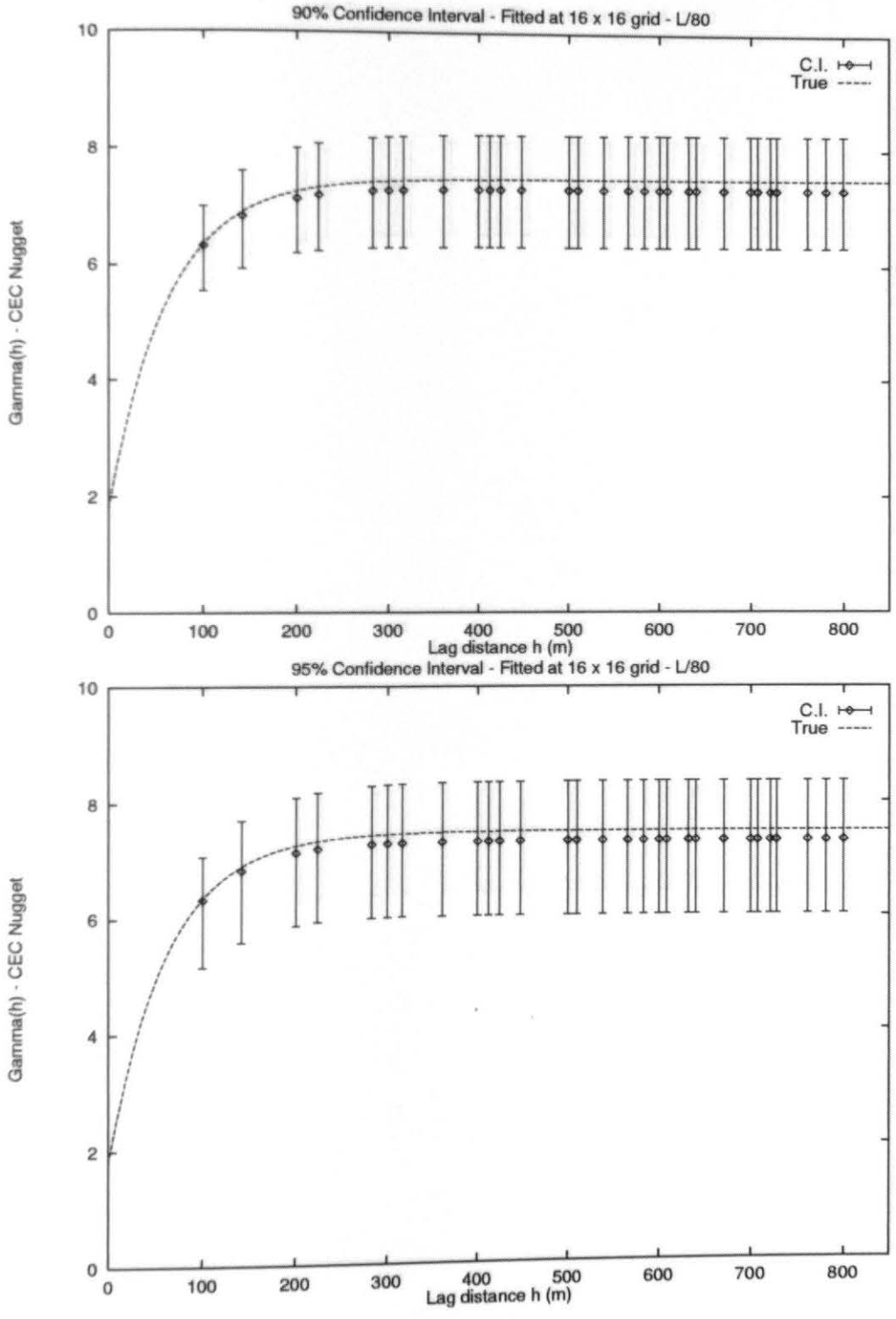


Figure C.11: Numerical Experiment 1 Nugget, 16x16 - L/80

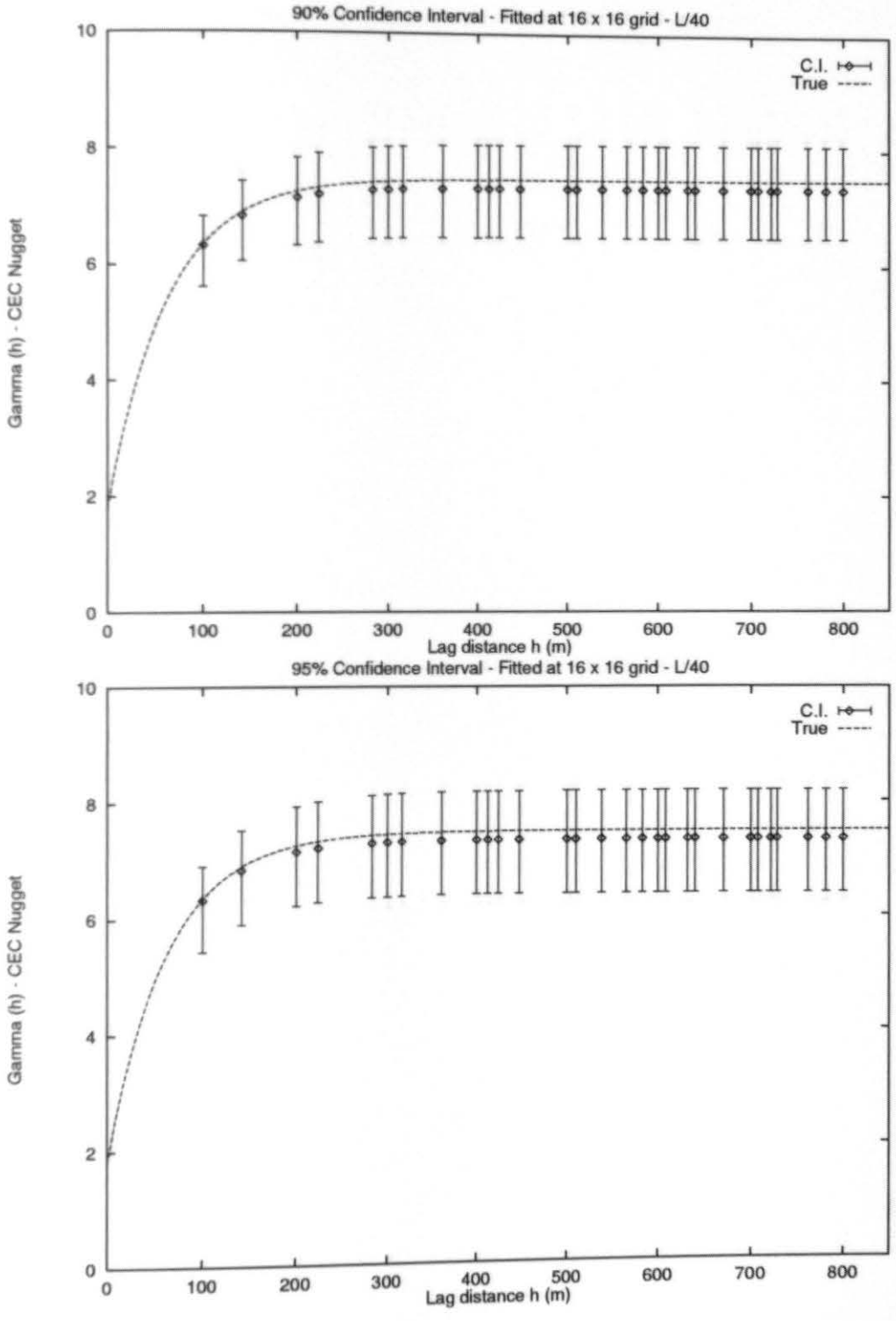


Figure C.12: Numerical Experiment 1 Nugget, 16x16 - L/40

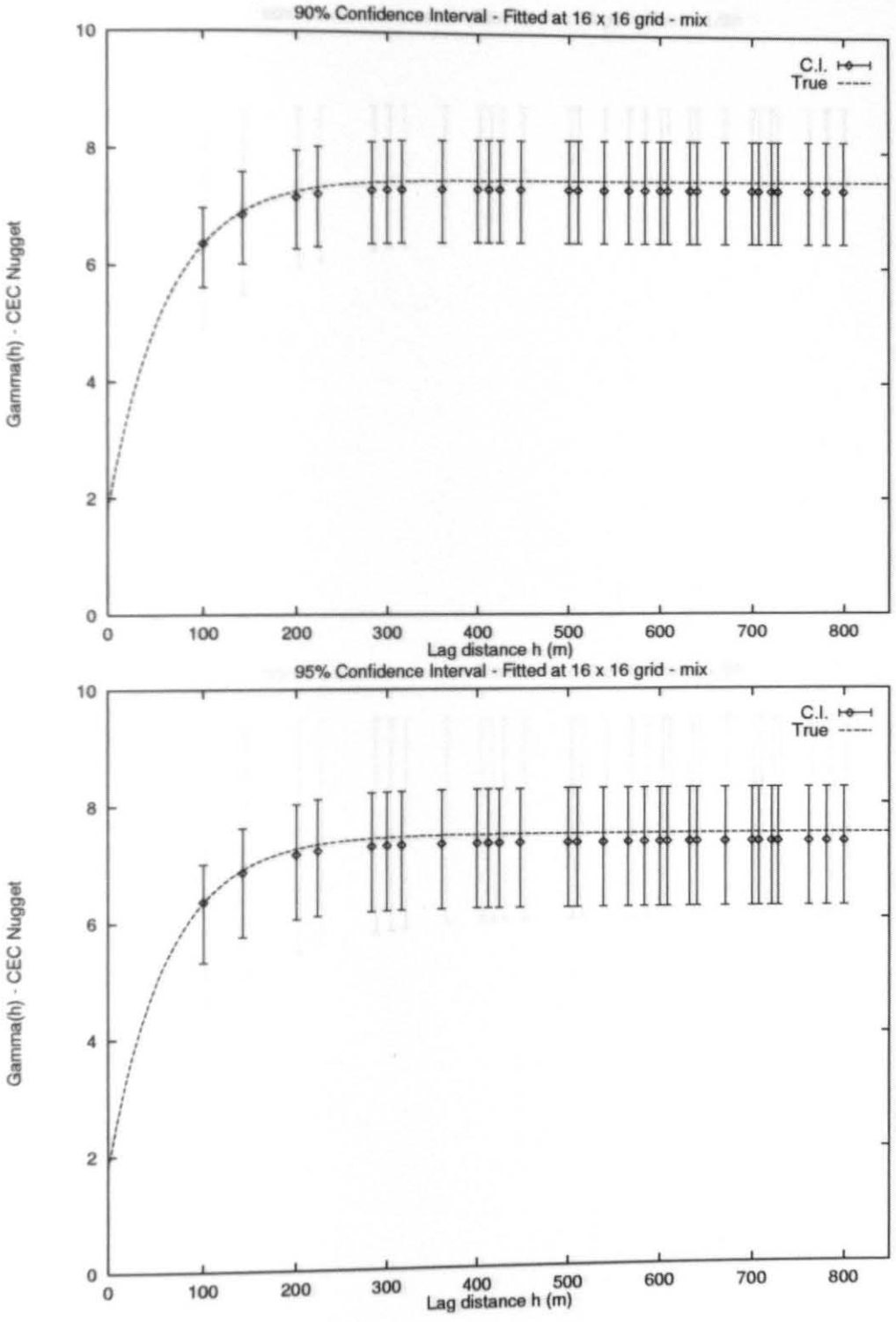


Figure C.13: Numerical Experiment 1 Nugget, 16x16 - mix

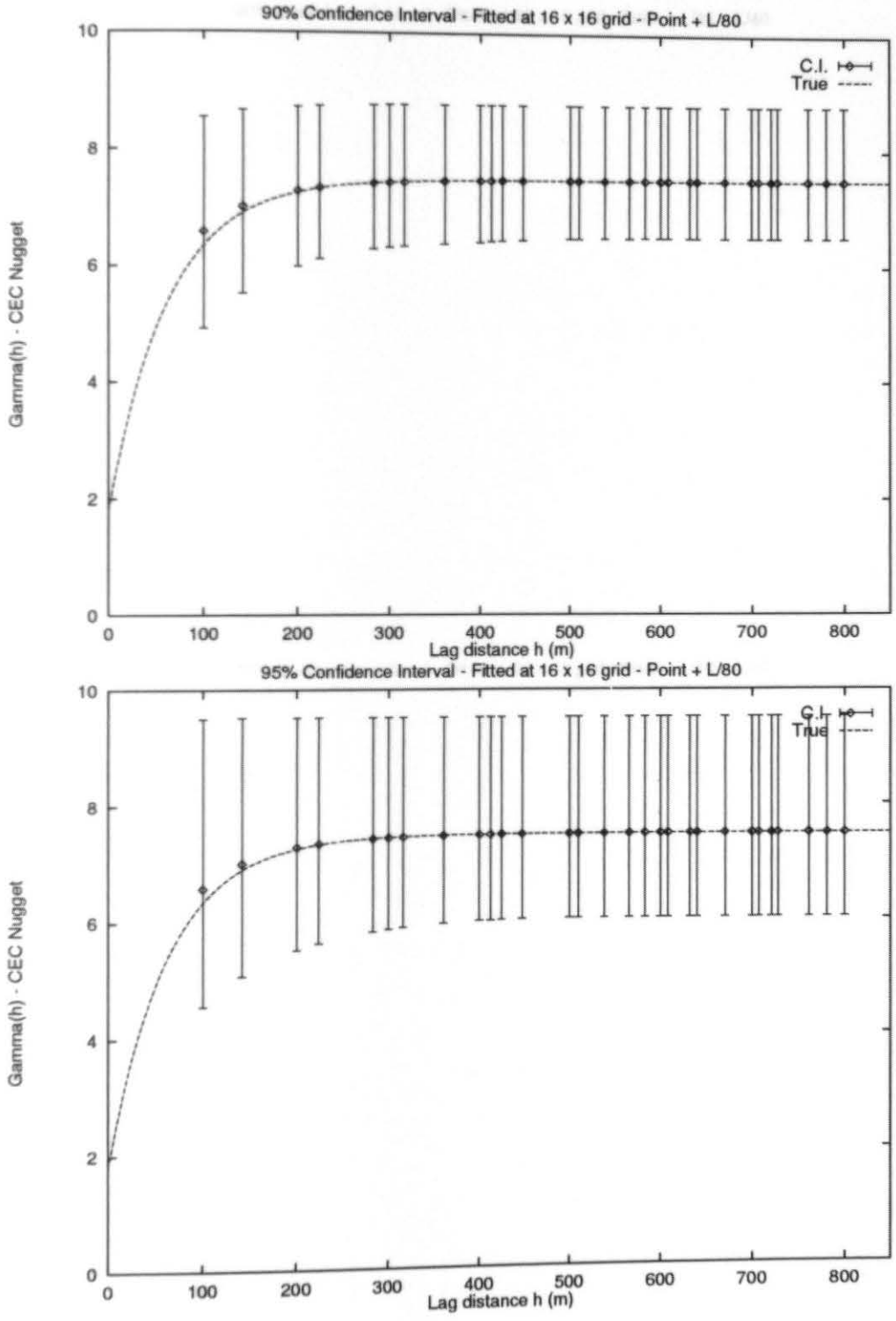


Figure C.14: Numerical Experiment 1 Nugget, 16x16 - point + L/80

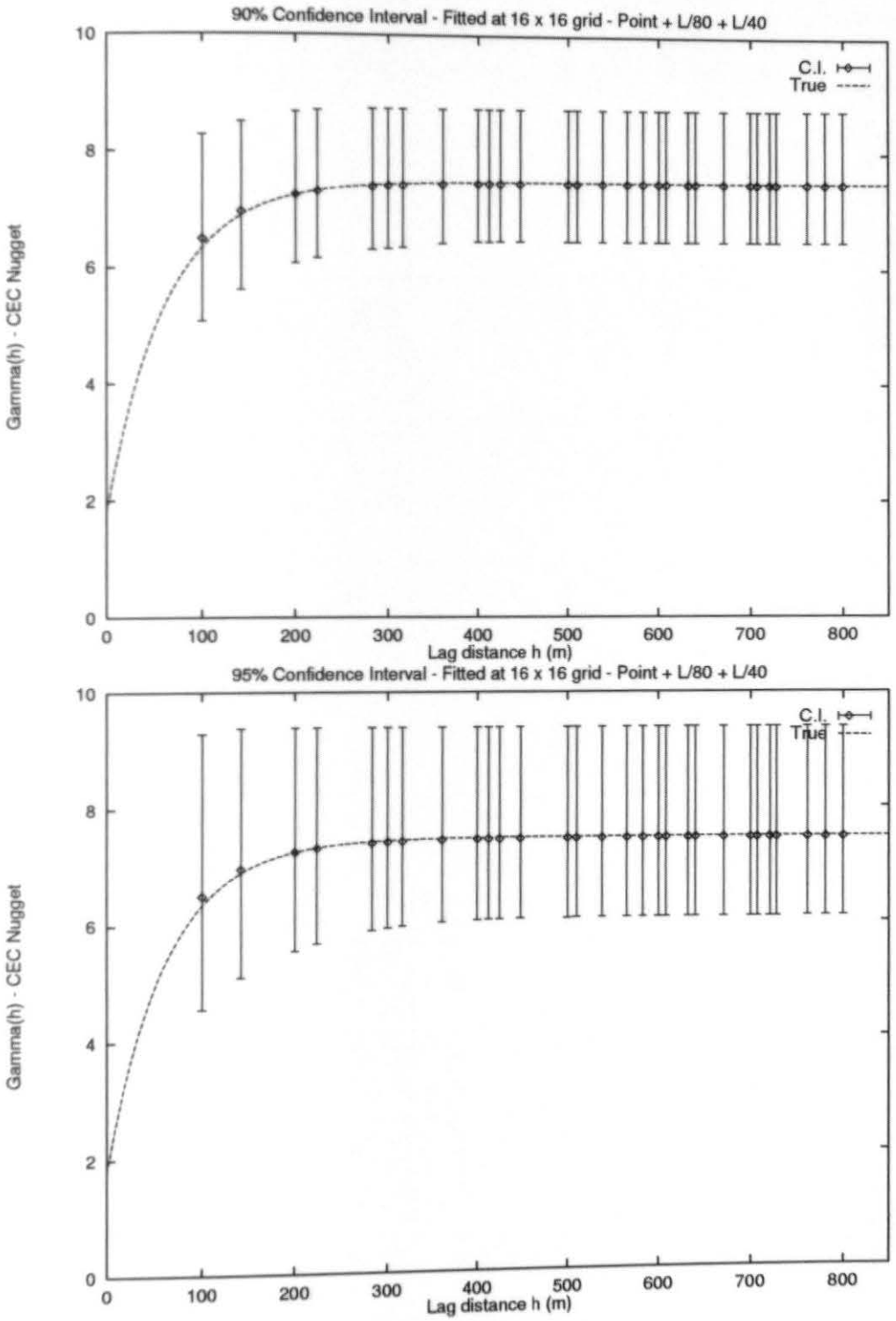


Figure C.15: Numerical Experiment 1 Nugget, 16x16 - point + L/80 + L/40

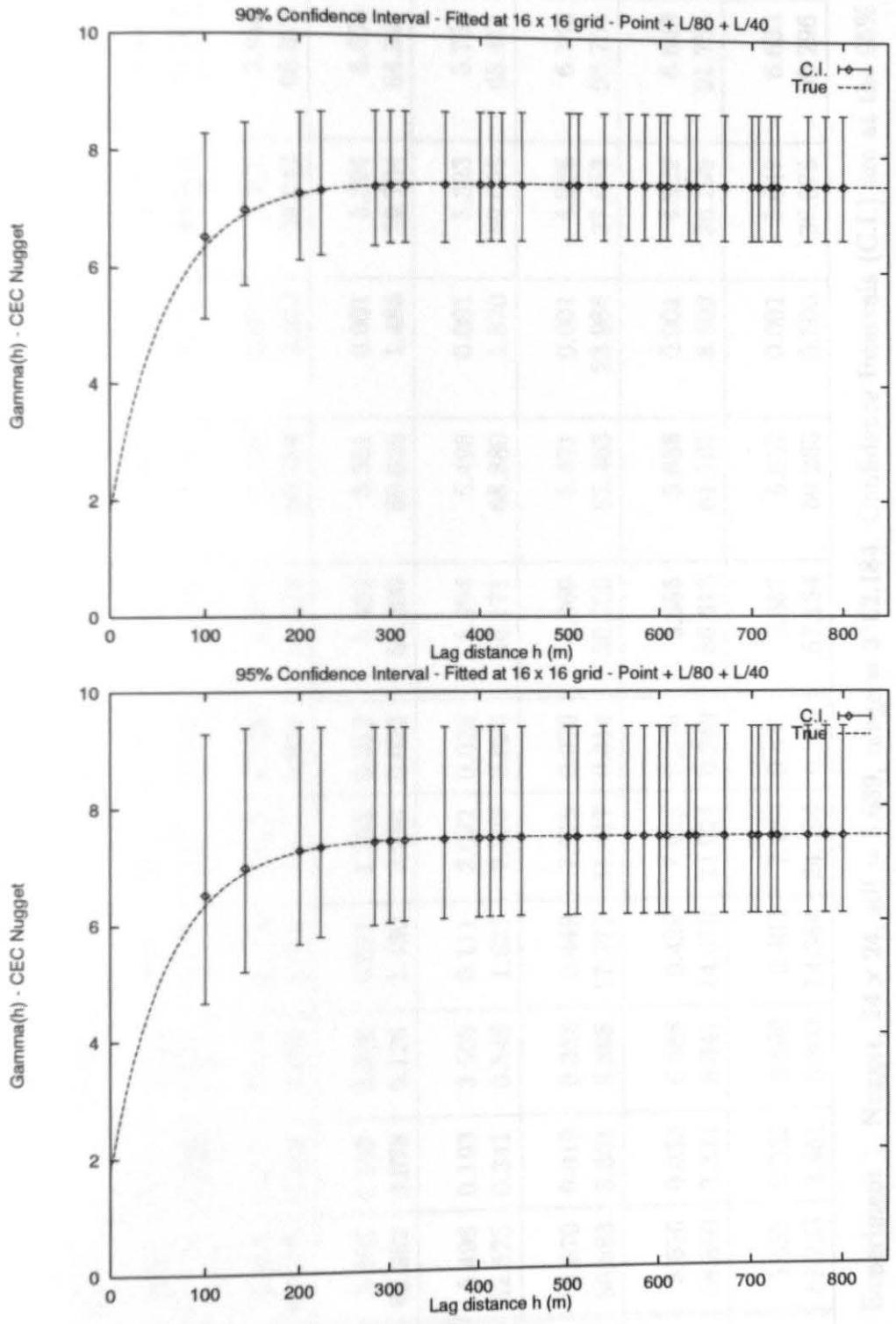


Figure C.16: Numerical Experiment 1 Nugget, 16x16 - point + L/80 + L/40 + mix

Parameter	Mean	Bias	%Bias	S.D.	%S.D.	C.V.	$\overline{\text{Low C.I.}}$	$\overline{\text{High C.I.}}$	S.D. C.I.	$\widehat{\text{Low C.I.}}$	$\widehat{\text{High C.I.}}$
Point:											
Sill	5.666	0.023	0.401	0.571	10.081	0.101	5.665	5.667	0.001	4.588	6.879
Range	61.215	0.969	1.583	6.723	10.983	0.110	47.414	75.016	14.759	49.884	75.904
L/80:											
Sill	5.492	0.197	3.587	0.159	2.899	0.029	5.491	5.493	0.001	5.226	5.843
Range	62.303	0.119	0.192	1.979	3.177	0.032	55.322	69.284	3.020	59.042	66.375
L/40:											
Sill	5.496	0.193	3.506	0.071	1.295	0.013	5.492	5.501	0.001	5.394	5.653
Range	62.262	0.078	0.126	1.420	2.280	0.023	55.500	69.025	1.488	59.784	64.841
Mixed:											
Sill	5.496	0.193	3.508	0.111	2.022	0.020	5.494	5.498	0.001	5.323	5.756
Range	62.525	0.341	0.546	1.621	2.593	0.026	56.171	68.880	1.870	59.698	65.461
Case A:											
Sill	5.670	0.019	0.333	0.449	7.919	0.079	5.669	5.671	0.001	4.968	6.748
Range	56.583	5.601	9.898	17.777	31.417	0.314	50.703	62.463	23.986	27.653	93.754
Case B:											
Sill	5.656	0.033	0.588	0.420	7.425	0.074	5.655	5.656	0.001	4.999	6.649
Range	58.860	3.324	5.647	14.670	24.924	0.249	56.613	61.107	8.502	36.199	91.755
Case C:											
Sill	5.657	0.032	0.566	0.403	7.120	0.071	5.657	5.657	0.001	5.047	6.633
Range	58.723	3.461	5.893	14.384	24.494	0.245	57.184	60.263	5.506	36.075	90.296

Table C.3: Experiment 1, Nugget, 24 x 24, sill = 5.689, range = 3*62.184. Confidence Intervals (C.I.) are at the 95% level

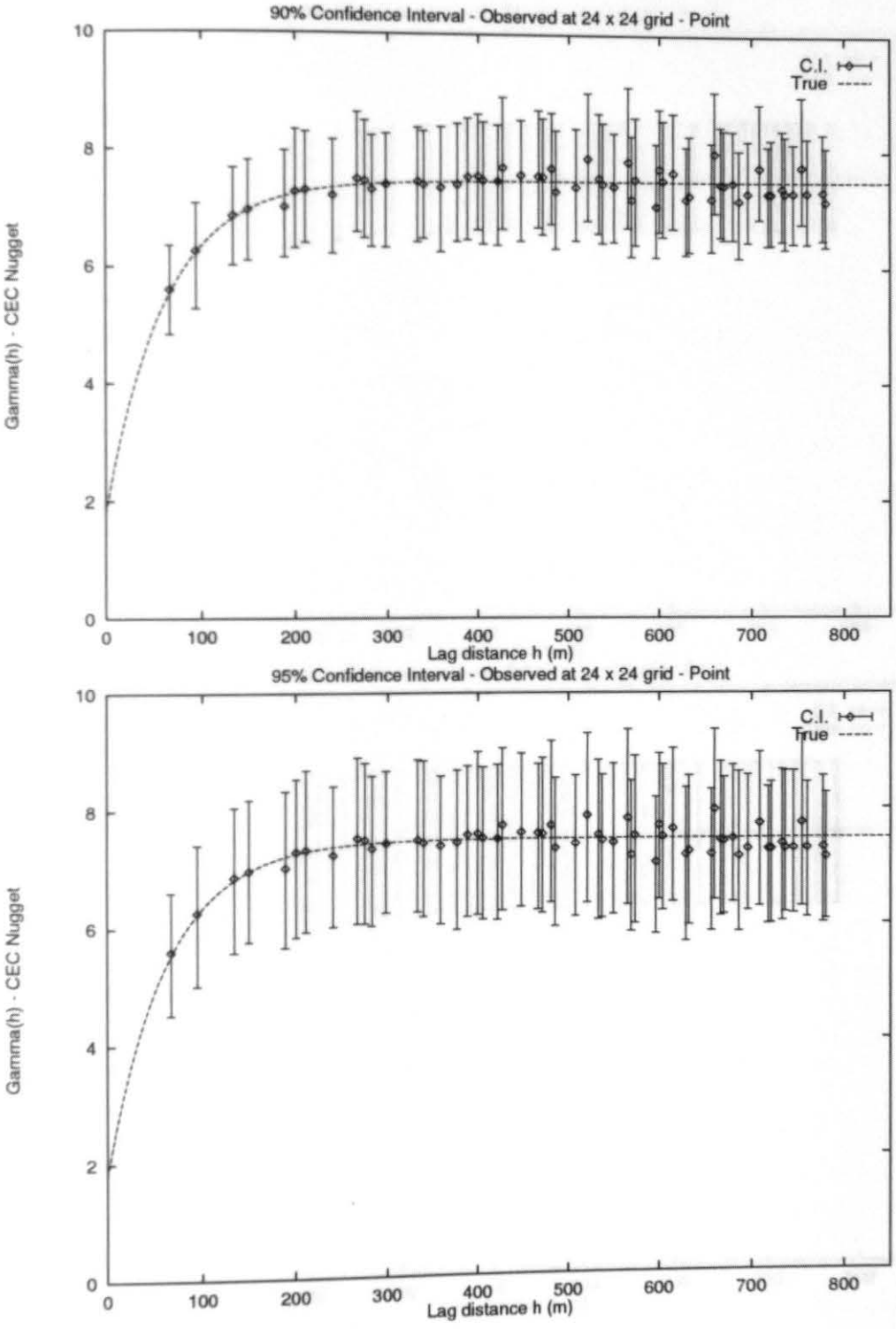


Figure C.17: Numerical Experiment 1 Nugget, 24x24 - Obs. Point

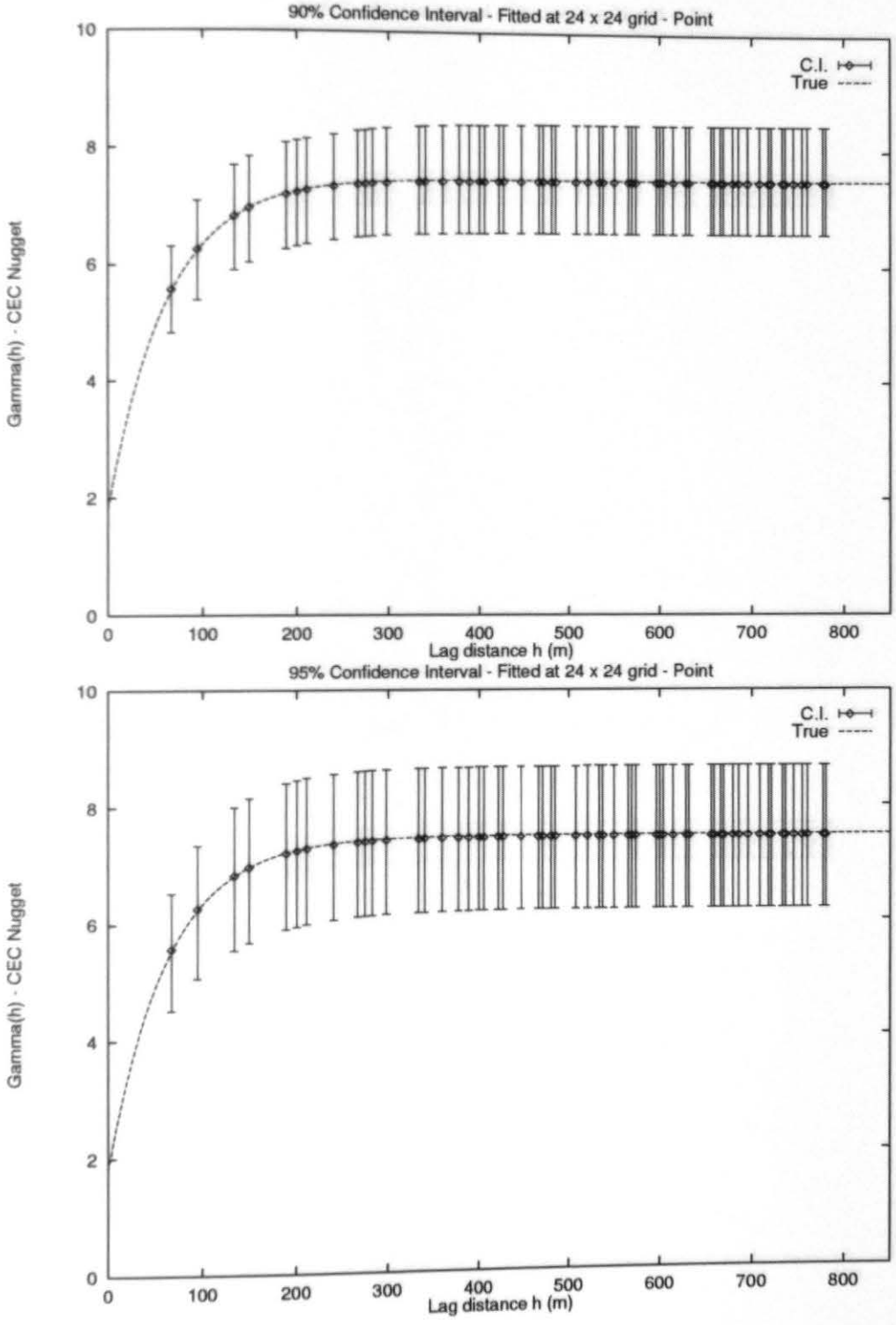


Figure C.18: Numerical Experiment 1 Nugget, 24x24 - Fit. Point

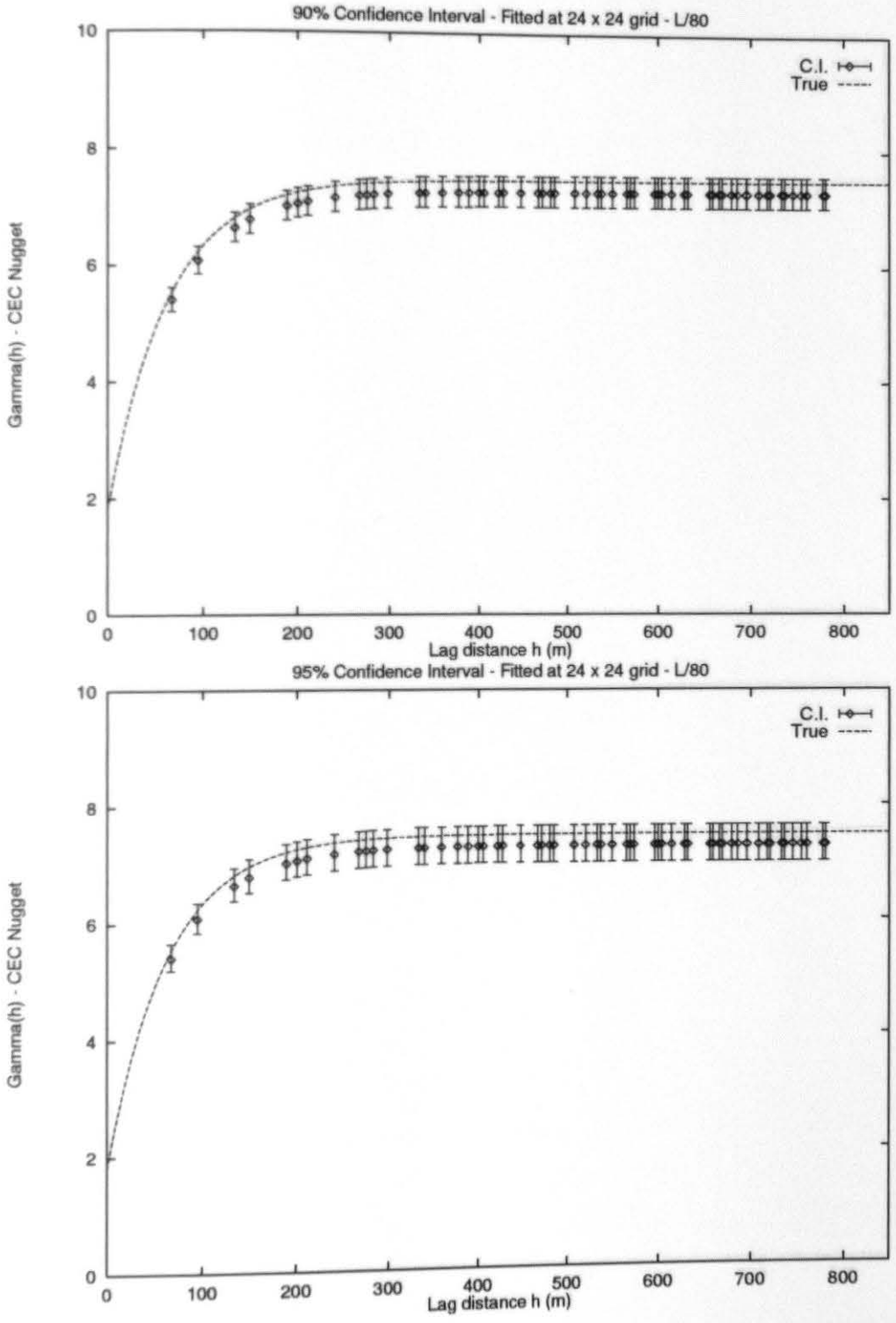


Figure C.19: Numerical Experiment 1 Nugget, 24x24 - L/80

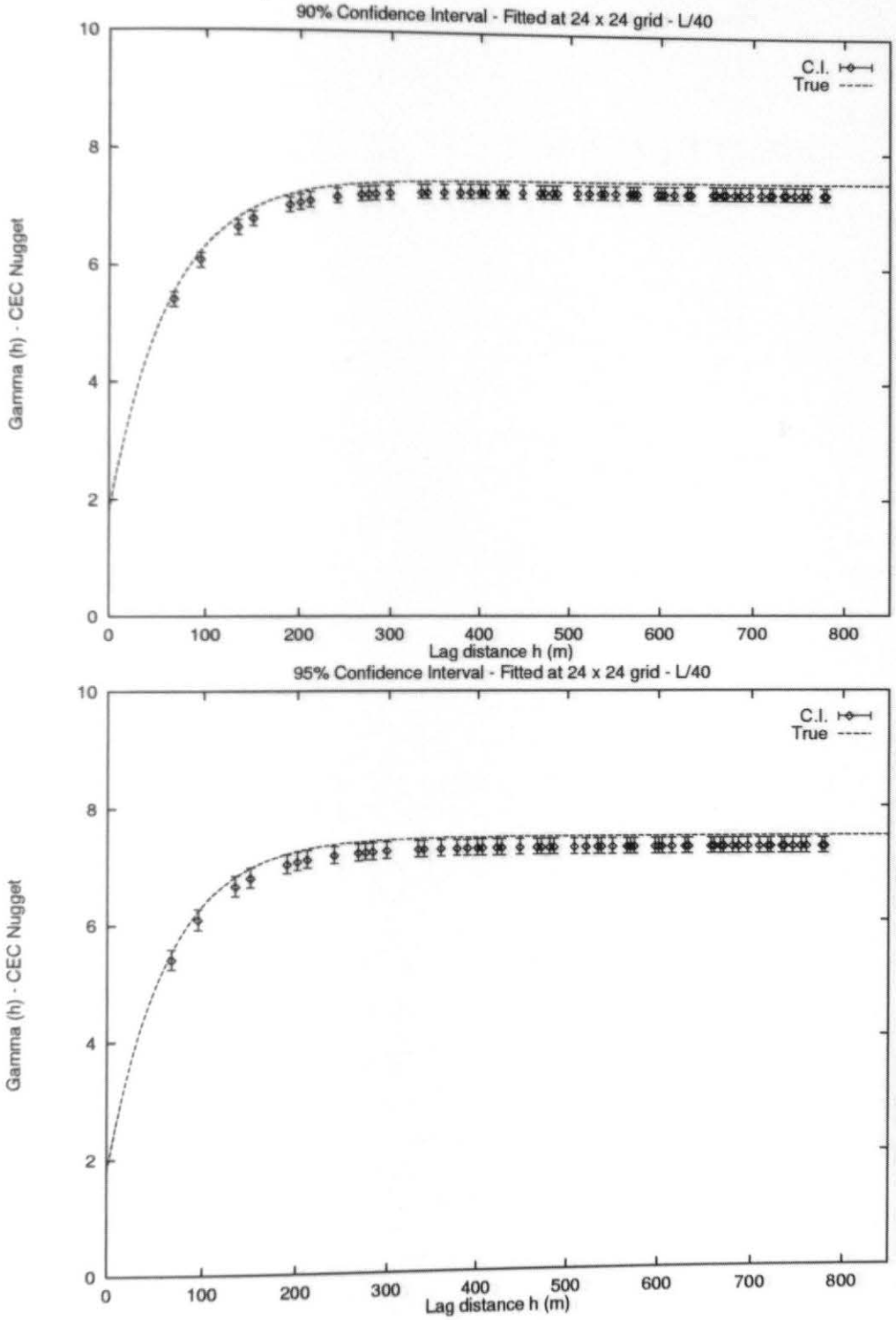


Figure C.20: Numerical Experiment 1 Nugget, 24x24 - L/40

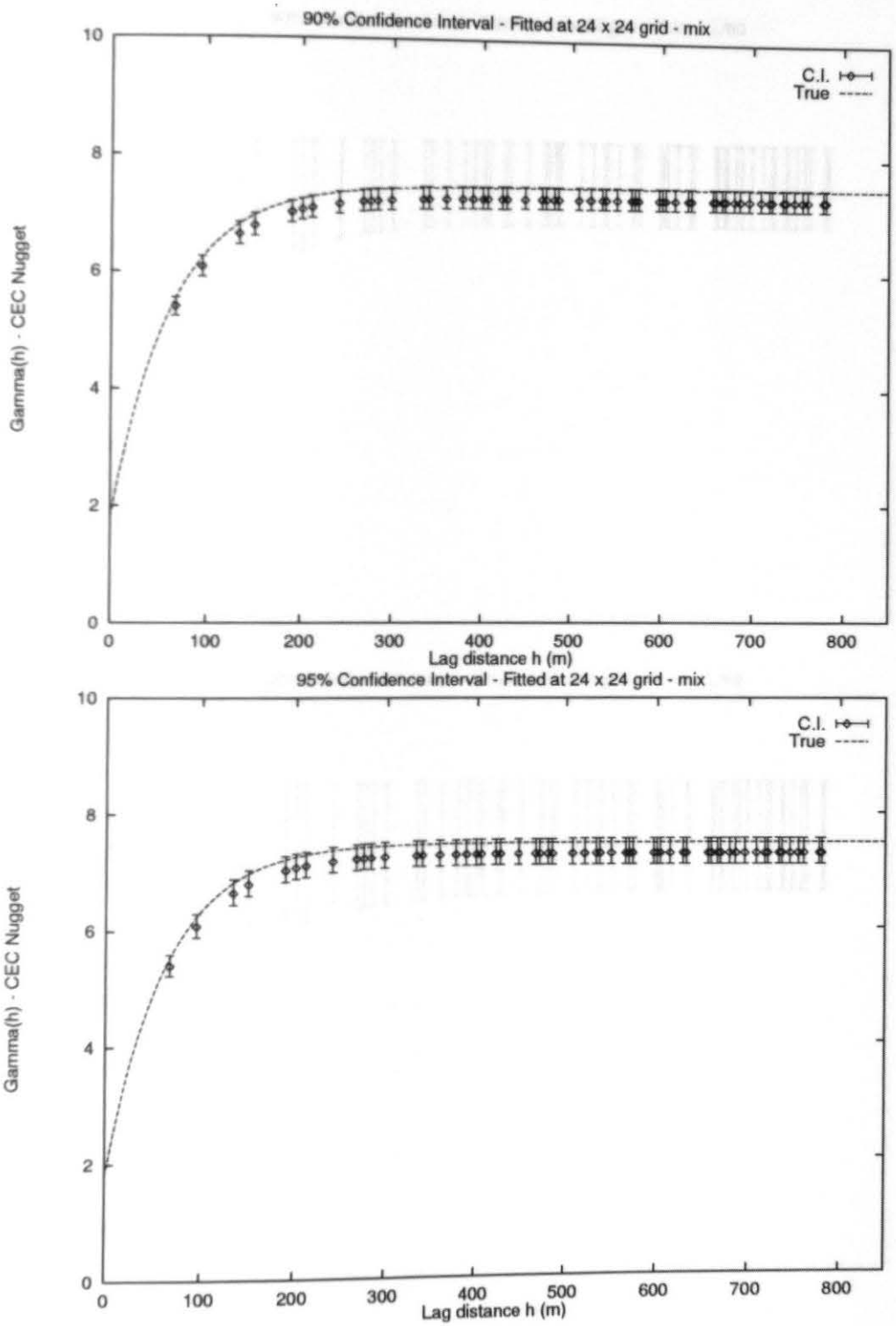


Figure C.21: Numerical Experiment 1 Nugget, 24x24 - mix

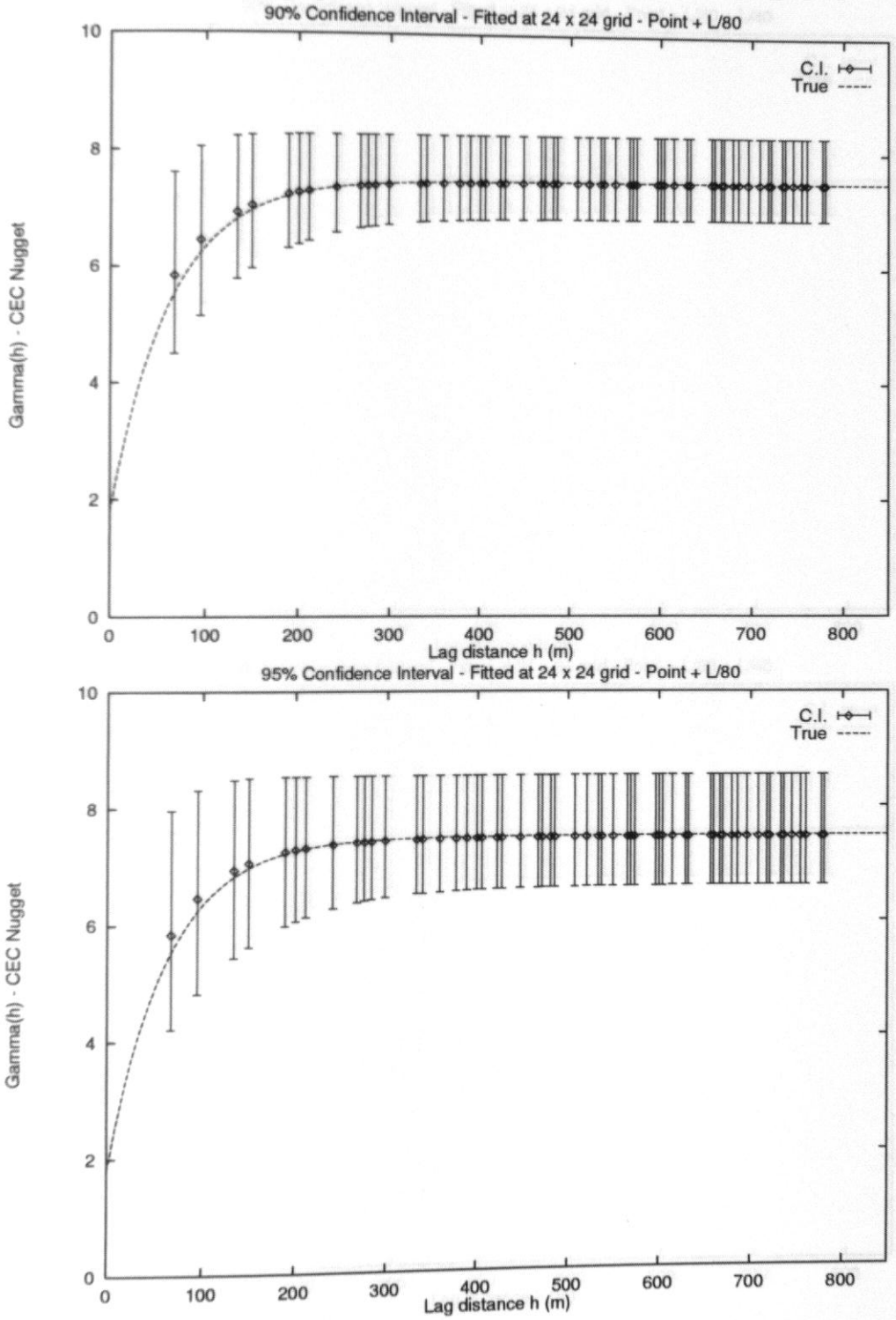


Figure C.22: Numerical Experiment 1 Nugget, 24x24 - point + L/80

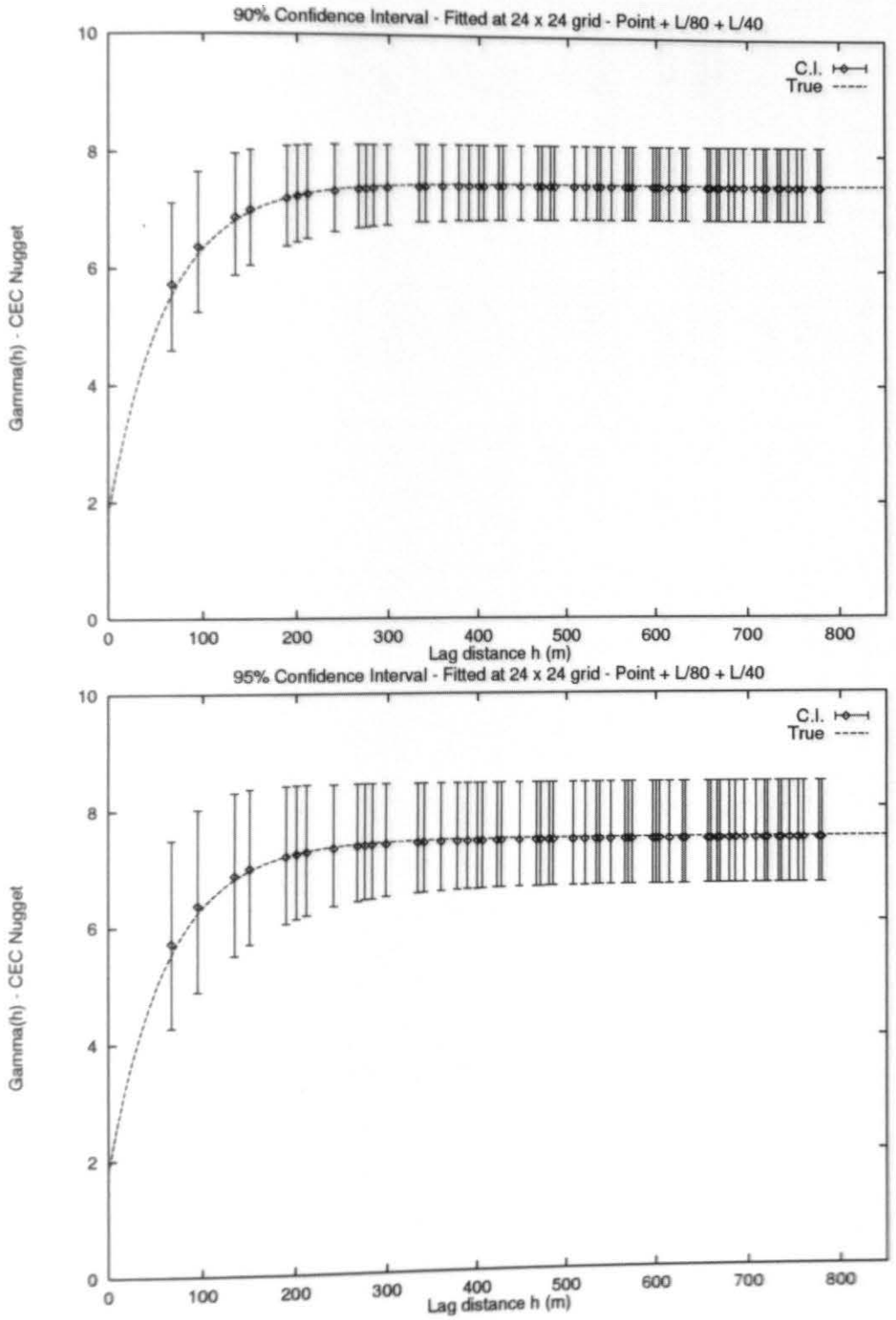


Figure C.23: Numerical Experiment 1 Nugget, 24x24 - point + L/80 + L/40

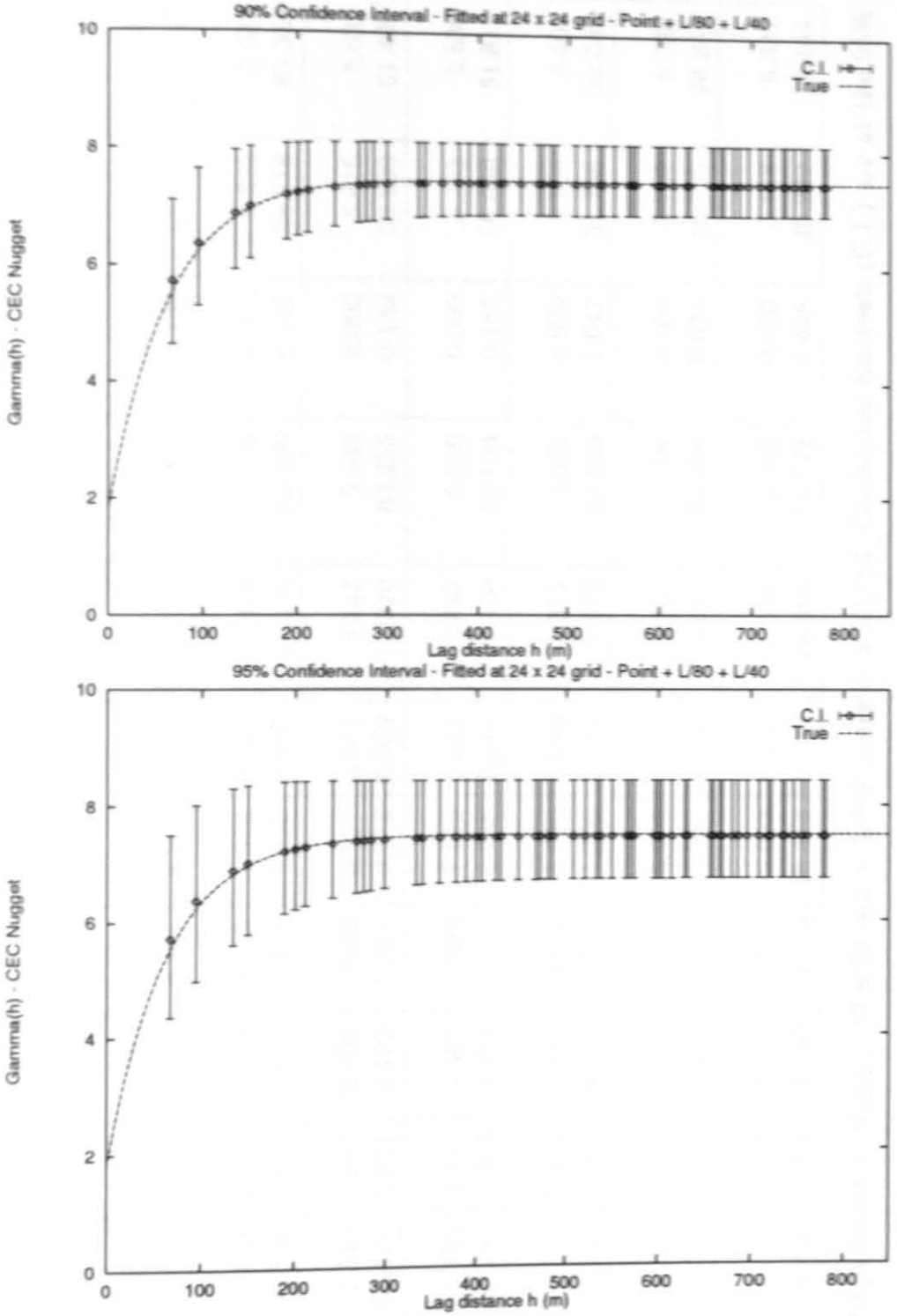


Figure C.24: Numerical Experiment 1 Nugget, 24x24 - point + L/80 + L/40 + mix

Parameter	Mean	Bias	%Bias	S.D.	%S.D.	C.V.	Low C.I.	High C.I.	S.D. C.I.	Low C.I.	High C.I.
Point:											
Sill	5.585	0.104	1.860	0.389	6.958	0.070	5.585	5.585	0.000	4.920	6.515
Range	61.098	1.086	1.778	4.441	7.268	0.073	56.712	65.484	3.138	52.979	70.727
L/80:											
Sill	5.459	0.230	4.218	0.083	1.519	0.015	5.458	5.459	0.000	5.323	5.637
Range	60.310	1.874	3.107	0.923	1.530	0.015	57.540	63.080	0.251	58.818	62.353
L/40:											
Sill	5.545	0.144	2.590	0.040	0.721	0.007	5.543	5.548	0.000	5.475	5.627
Range	60.312	1.872	3.103	0.557	0.924	0.009	57.370	63.255	0.139	59.299	61.427
Mixed:											
Sill	5.498	0.191	3.465	0.063	1.139	0.011	5.497	5.500	0.000	5.393	5.630
Range	60.280	1.904	3.158	0.747	1.238	0.012	57.456	63.104	0.186	59.040	61.815
Case A:											
Sill	5.571	0.118	2.112	0.311	5.590	0.056	5.571	5.572	0.000	5.104	6.404
Range	56.752	5.432	9.571	12.120	21.356	0.214	55.620	57.884	1.647	31.736	78.589
Case B:											
Sill	5.566	0.123	2.215	0.290	5.215	0.052	5.566	5.566	0.000	5.122	6.338
Range	59.589	2.595	4.355	9.880	16.580	0.166	59.098	60.080	0.632	40.505	78.247
Case C:											
Sill	5.566	0.123	2.211	0.279	5.007	0.050	5.566	5.566	0.000	5.148	6.322
Range	59.358	2.826	4.760	9.743	16.413	0.164	58.995	59.722	0.458	40.239	77.464

Table C.4: Experiment 1, Nugget, 32 x 32, sill = 5.689, range = 3*62.184. Confidence Intervals (C.I.) are at the 95% level

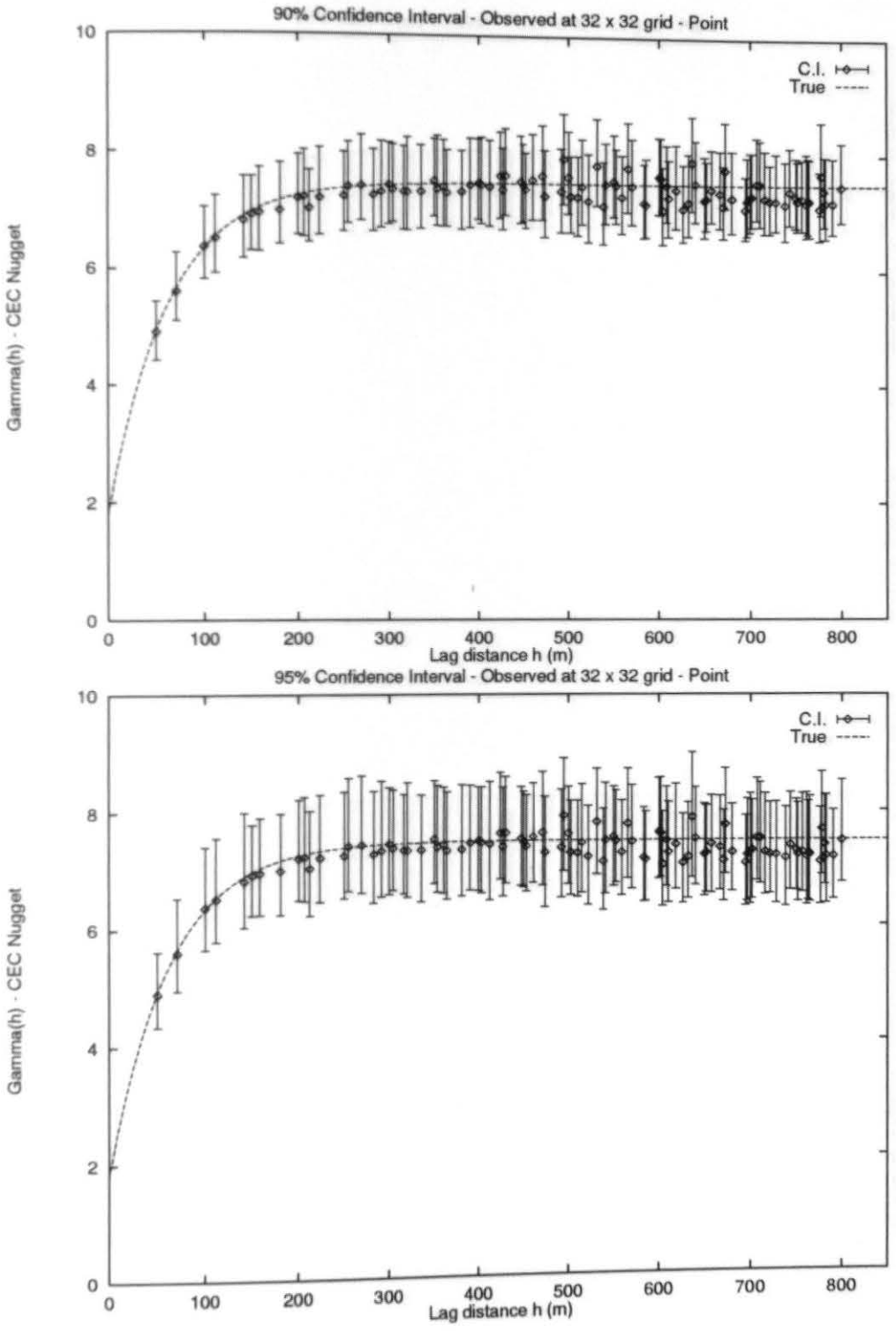


Figure C.25: Numerical Experiment 1 Nugget, 32x32 - Obs. Point

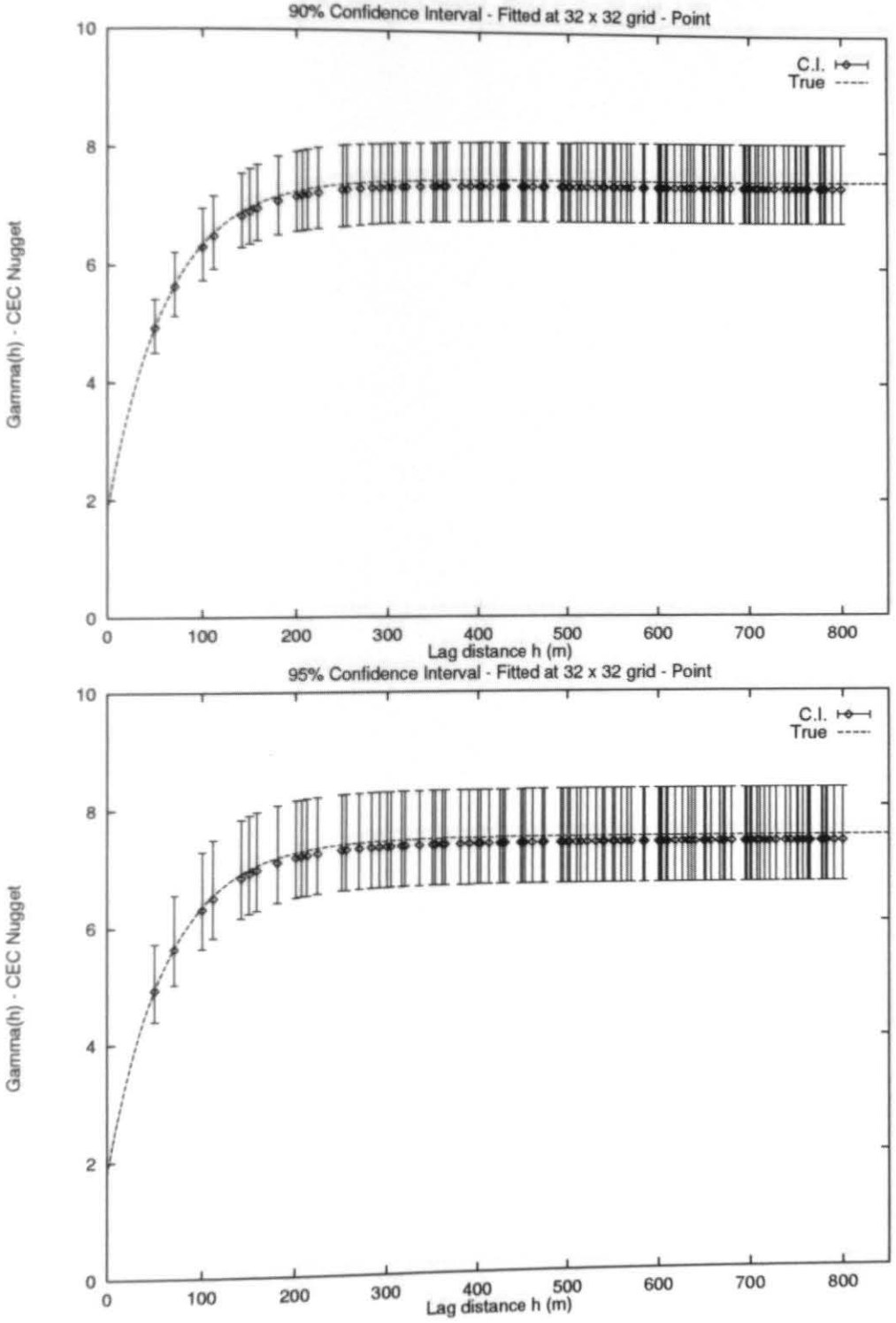


Figure C.26: Numerical Experiment 1 Nugget, 32x32 - Fit. Point

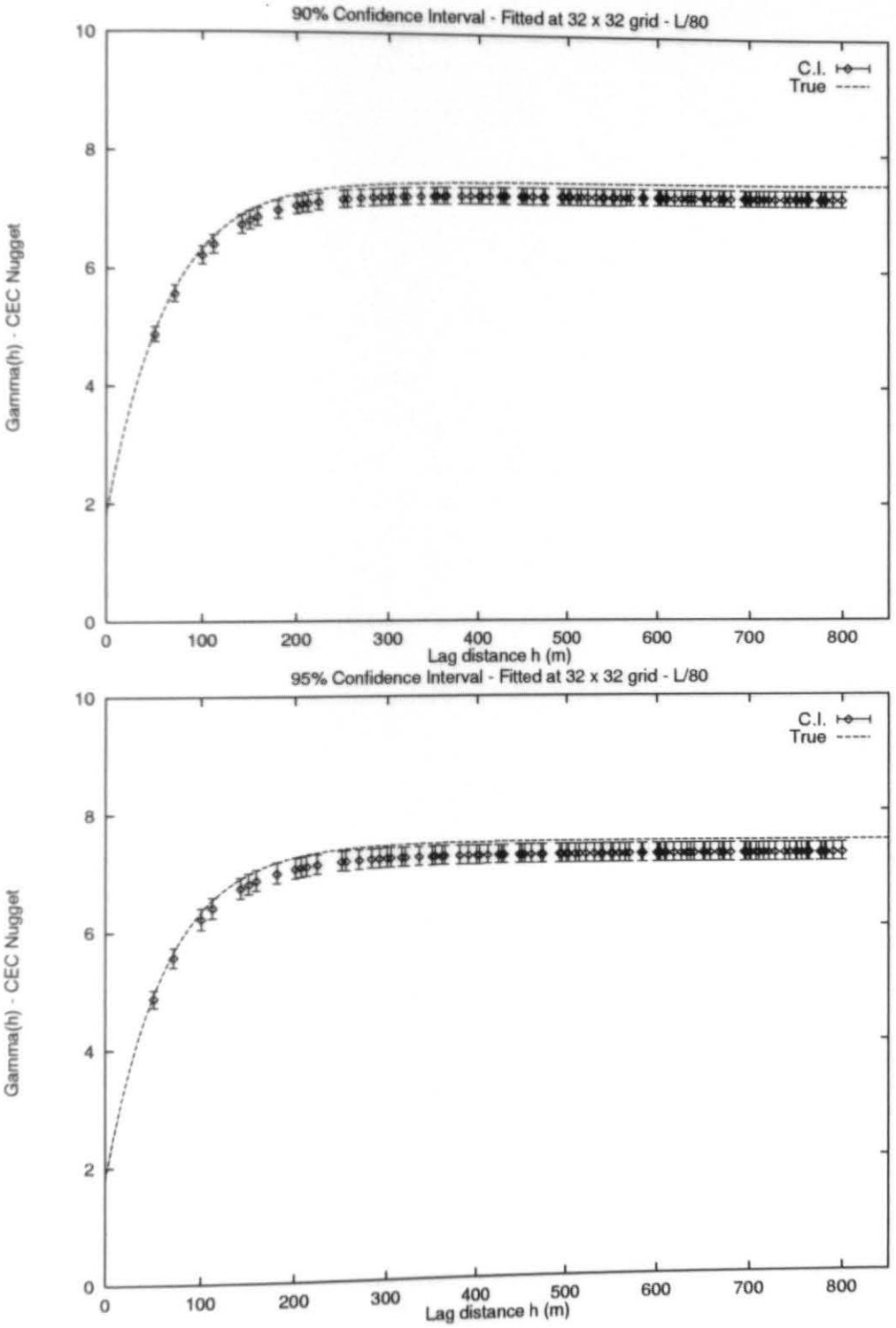


Figure C.27: Numerical Experiment 1 Nugget, 32x32 - L/80

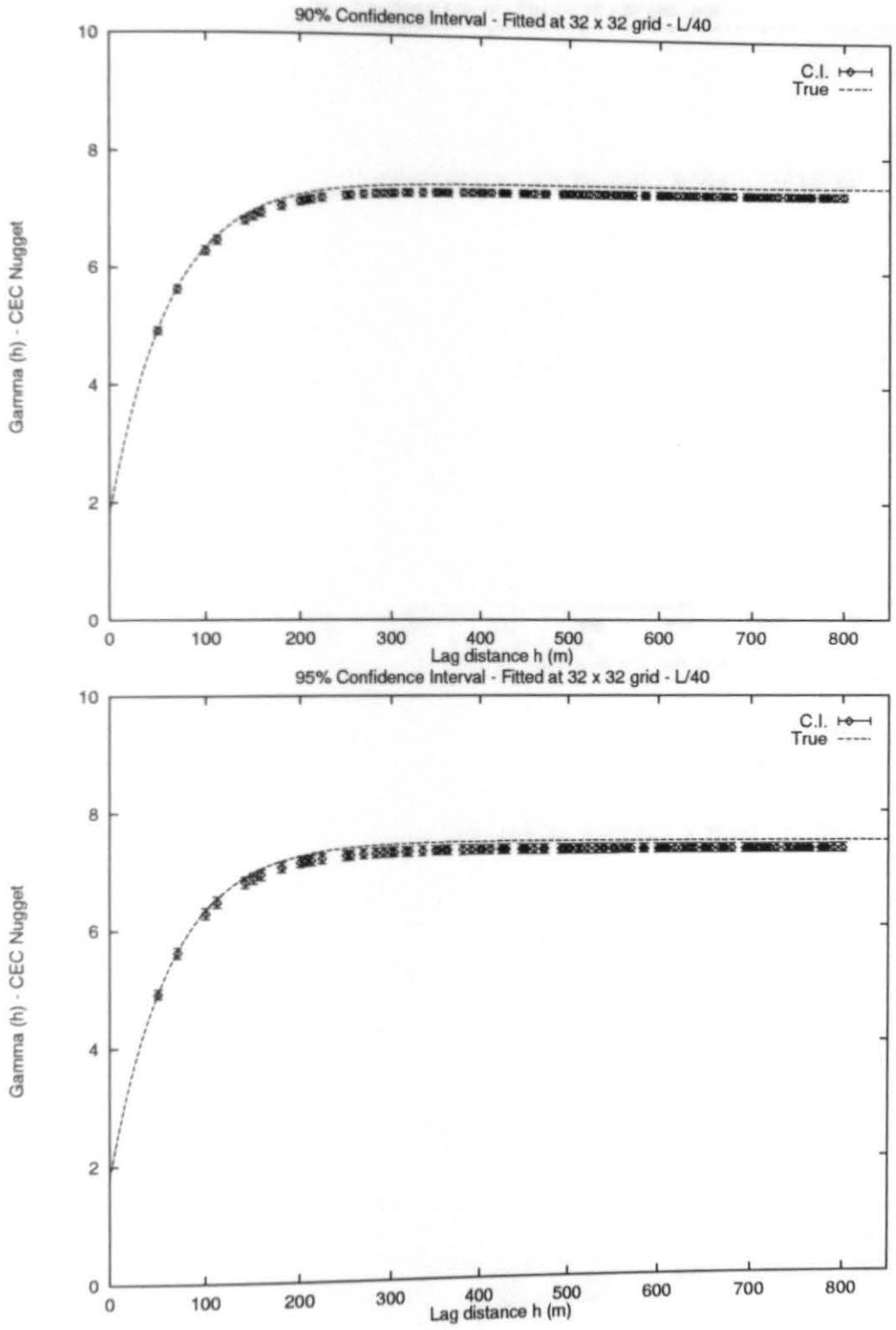


Figure C.28: Numerical Experiment 1 Nugget, 32x32 - L/40

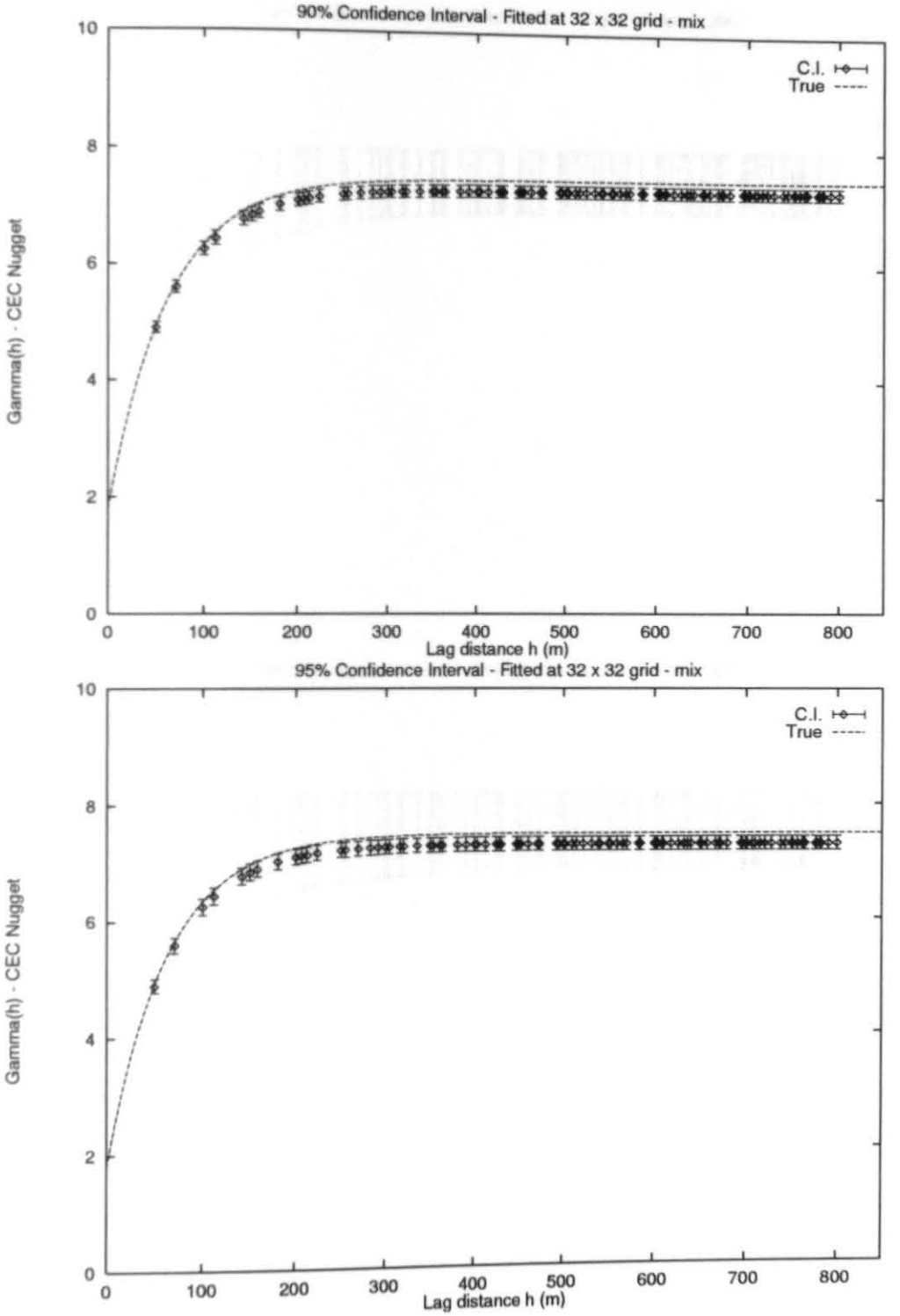


Figure C.29: Numerical Experiment 1 Nugget, 32x32 - mix

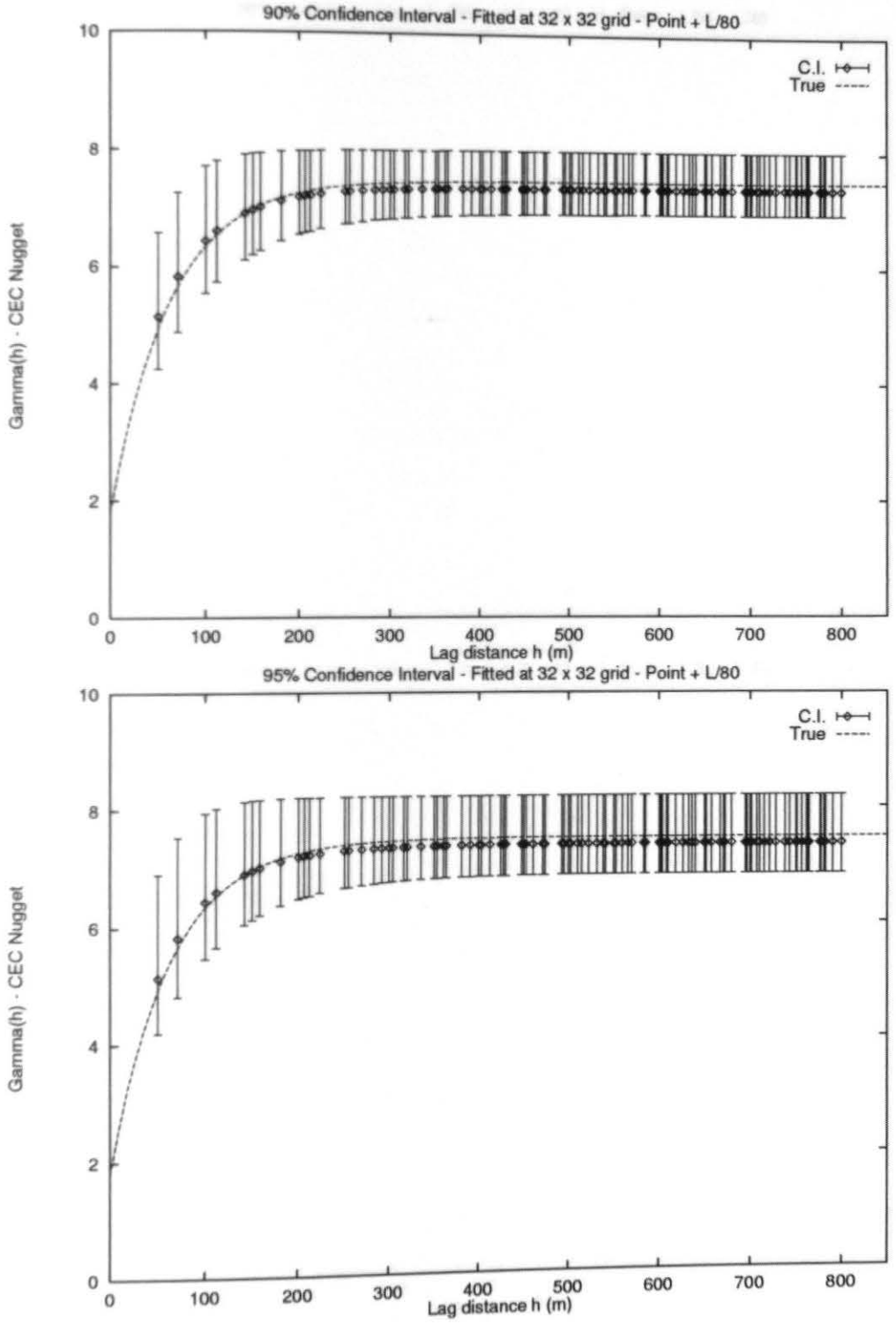


Figure C.30: Numerical Experiment 1 Nugget, 32x32 - point + L/80

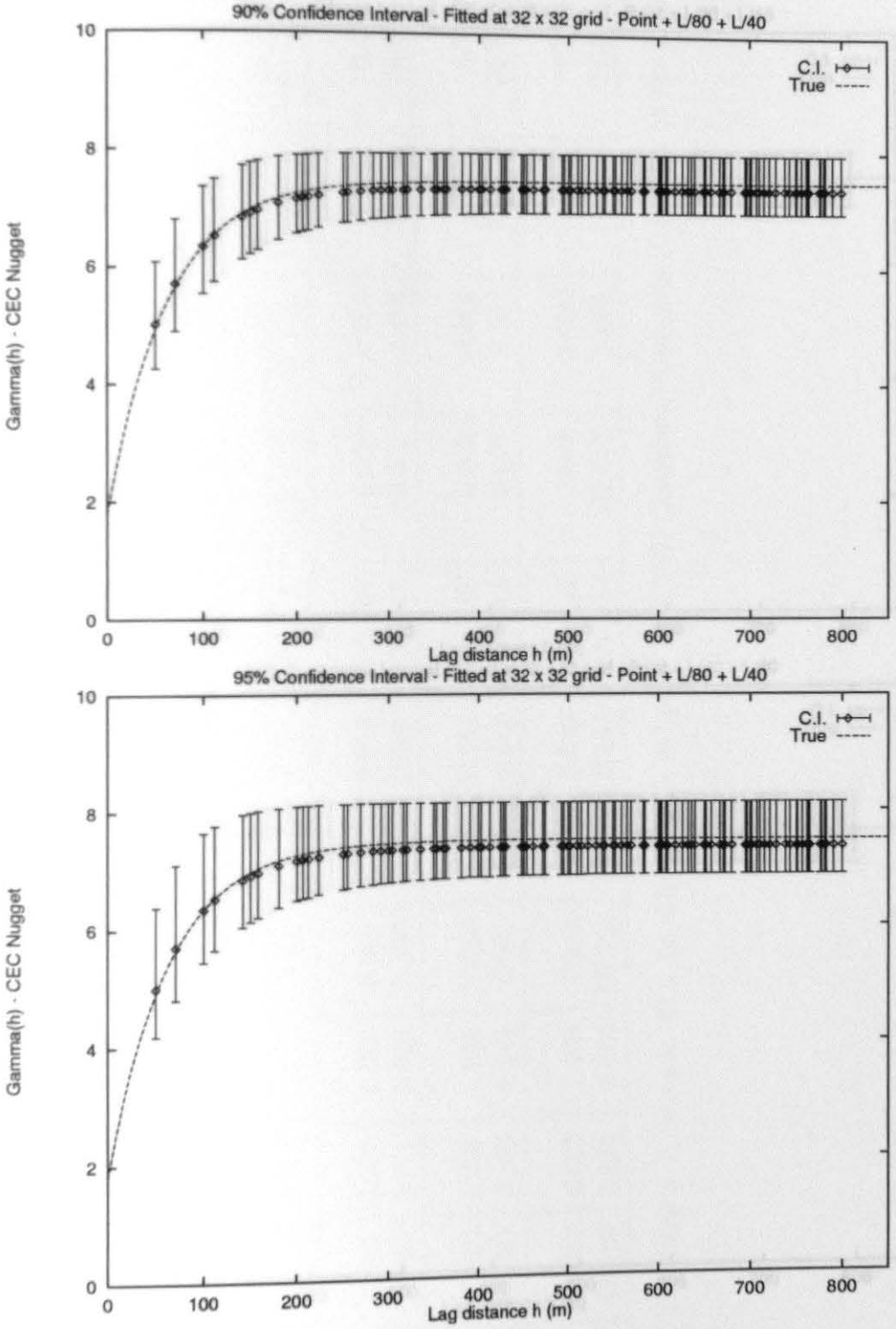


Figure C.31: Numerical Experiment 1 Nugget, 32x32 - point + L/80 + L/40

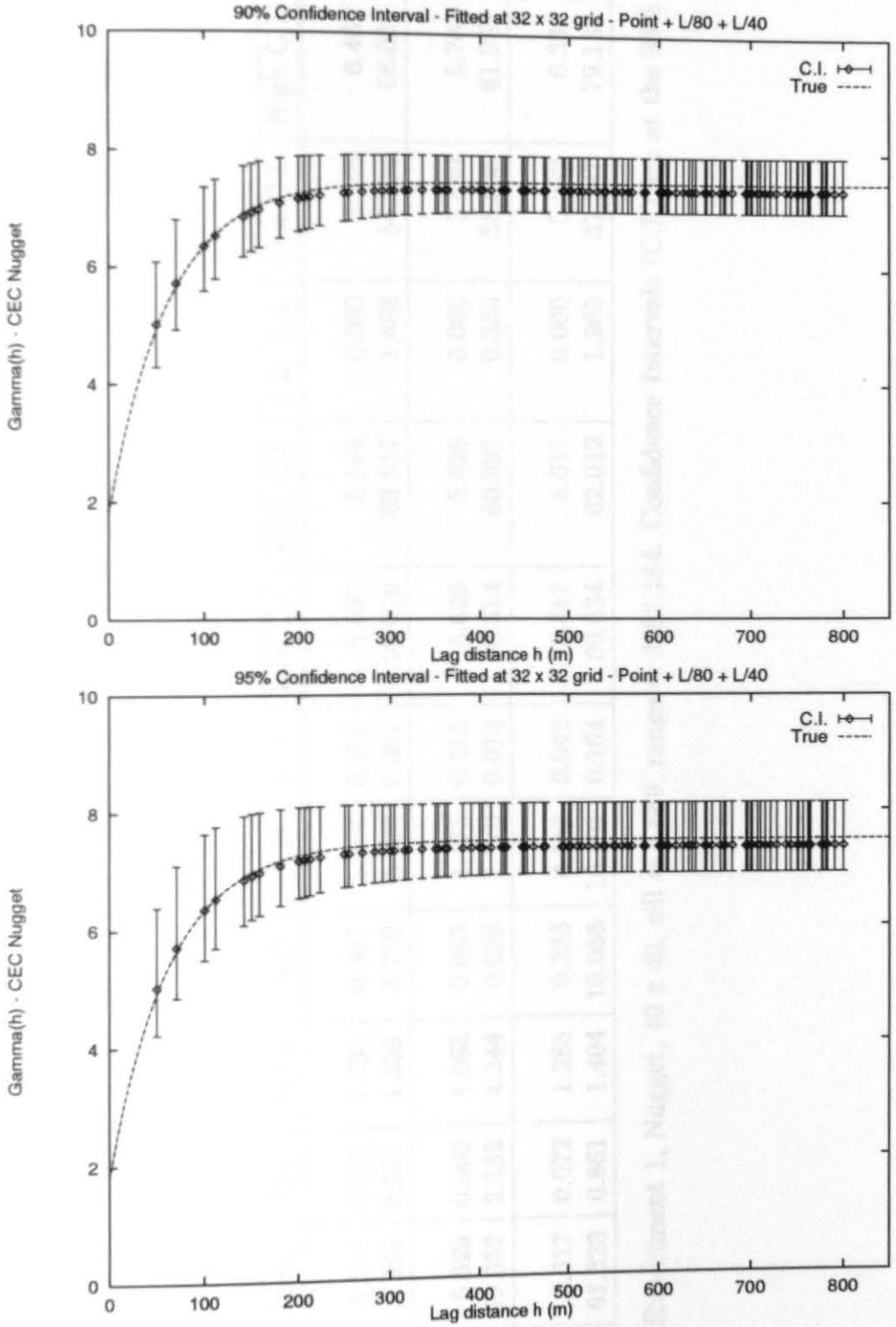


Figure C.32: Numerical Experiment 1 Nugget, 32x32 - point + L/80 + L/40 + mix

Parameter	Mean	Bias	%Bias	S.D.	%S.D.	C.V.	$\overline{\text{Low C.I.}}$	$\overline{\text{High C.I.}}$	S.D. C.I.	$\widehat{\text{Low C.I.}}$	$\widehat{\text{High C.I.}}$
Point:											
Sill	5.597	0.092	1.635	0.307	5.484	0.055	5.597	5.598	0.000	5.146	6.469
Range	61.263	0.921	1.503	3.750	6.122	0.061	59.010	63.517	1.488	54.234	68.037
L/80:											
Sill	5.629	0.060	1.065	0.083	1.475	0.015	5.629	5.629	0.000	5.449	5.763
Range	59.652	2.532	4.244	0.698	1.170	0.012	58.314	60.991	0.350	58.565	61.257
Case A:											
Sill	5.617	0.072	1.285	0.233	4.153	0.042	5.617	5.617	0.000	5.292	6.337
Range	61.323	0.861	1.404	10.058	16.401	0.164	60.634	62.012	1.265	37.247	79.182

Table C.5: Experiment 1, Nugget, 40 x 40, sill = 5.689, range = 3*62.184. Confidence Intervals (C.I.) are at the 95% level

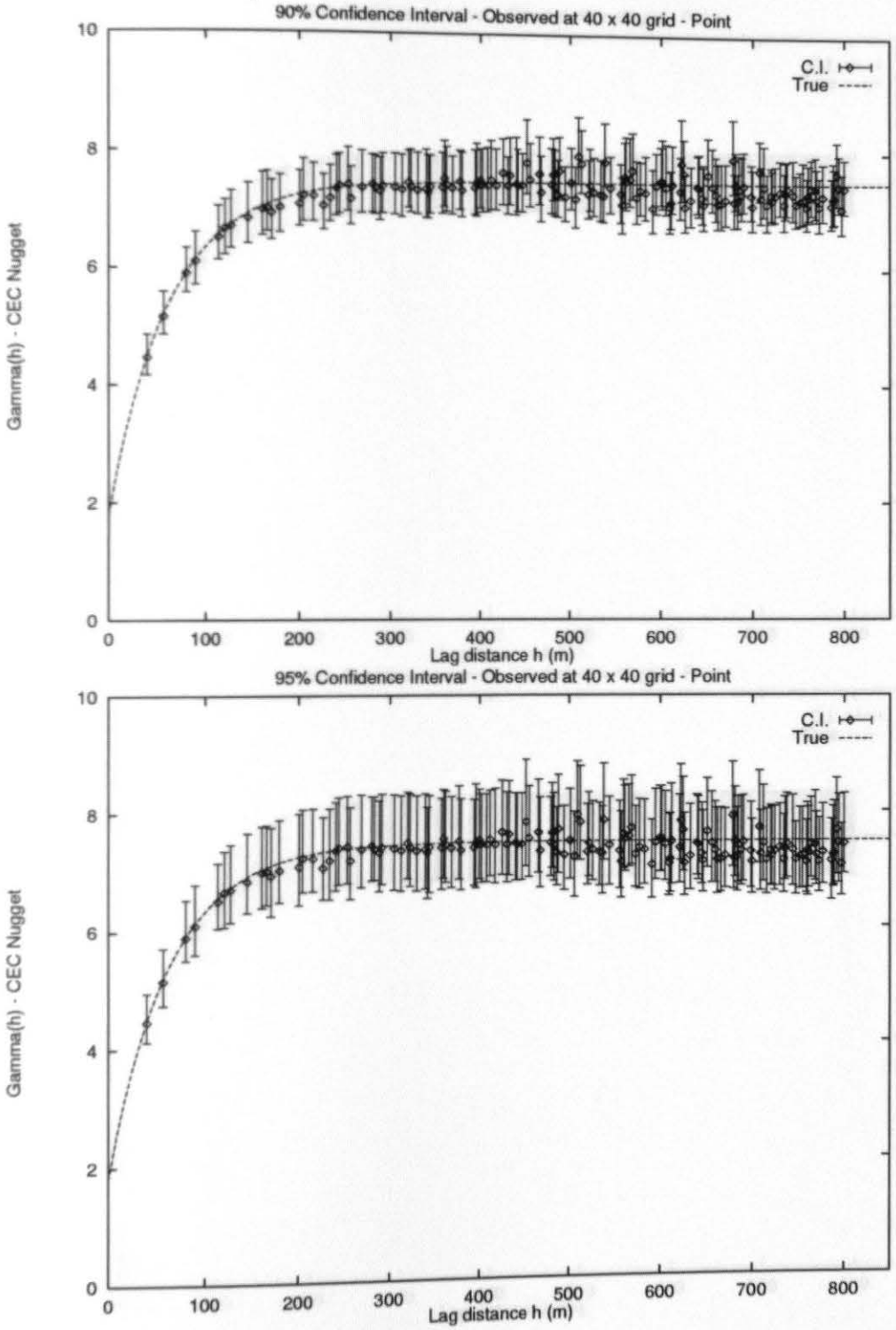


Figure C.33: Numerical Experiment 1 Nugget, 40x40 - Obs. point

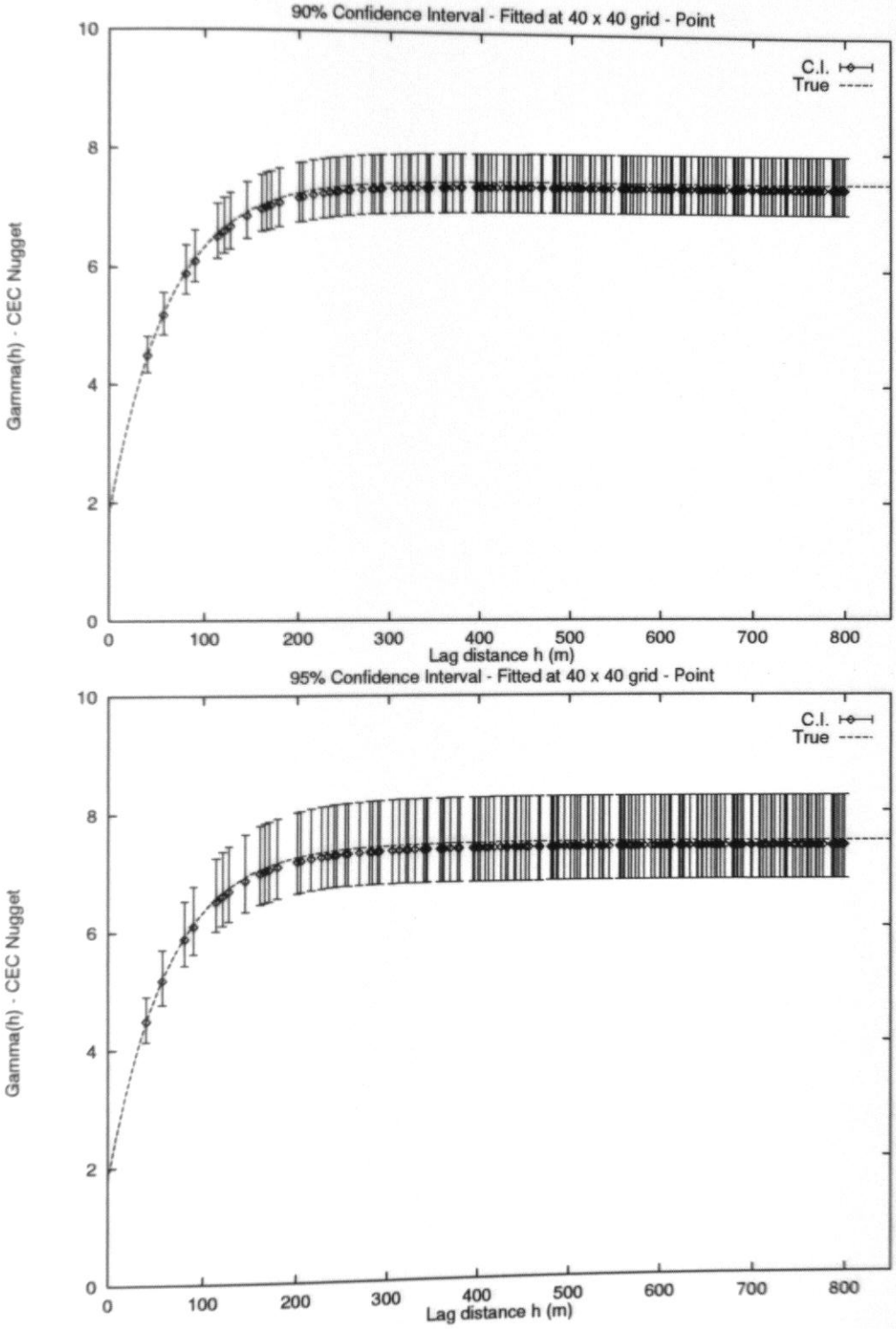


Figure C.34: Numerical Experiment 1 Nugget, 40x40 - Fit. point

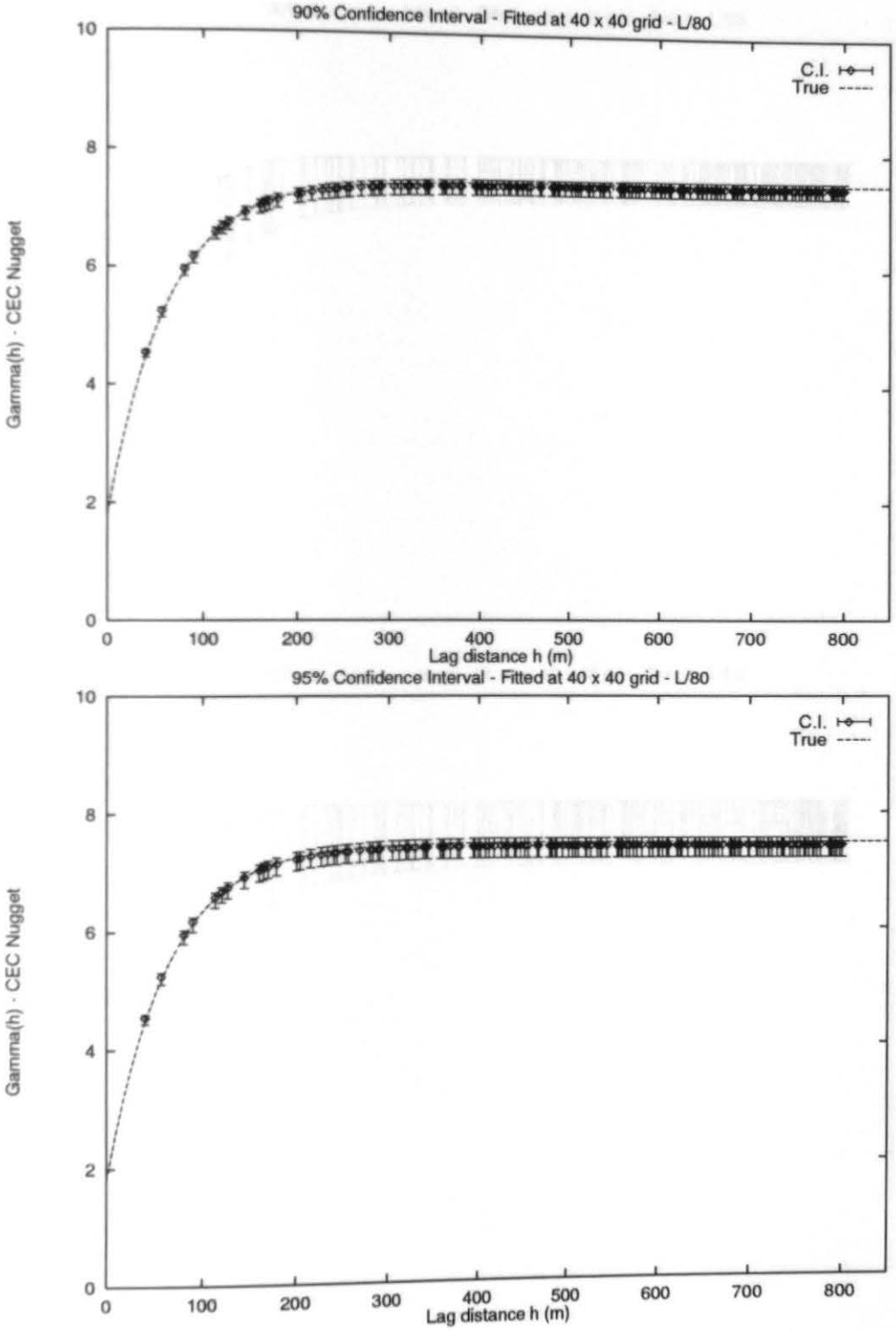


Figure C.35: Numerical Experiment 1 Nugget, 40x40 - L/80

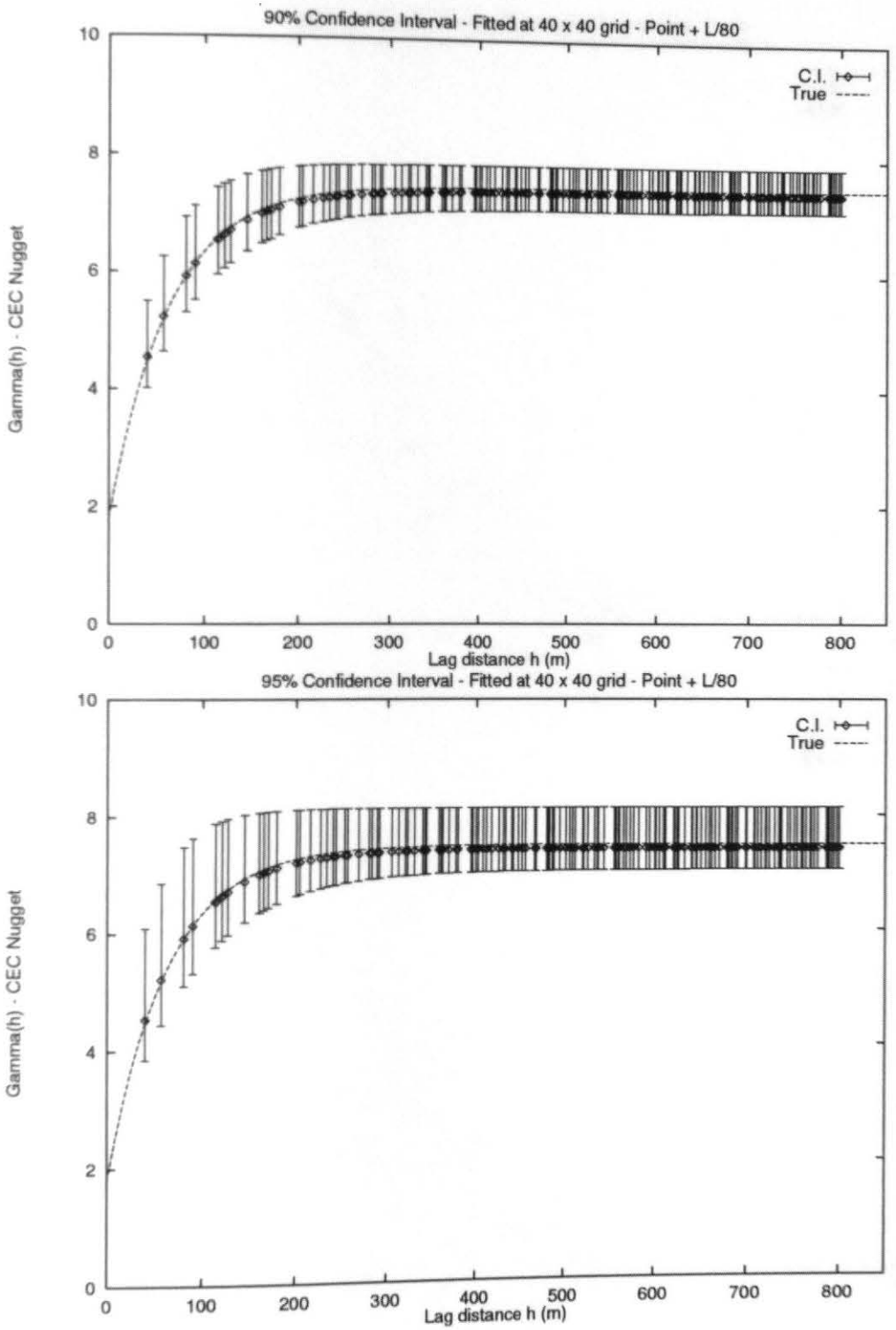


Figure C.36: Numerical Experiment 1 Nugget, 40x40 - point + L/80

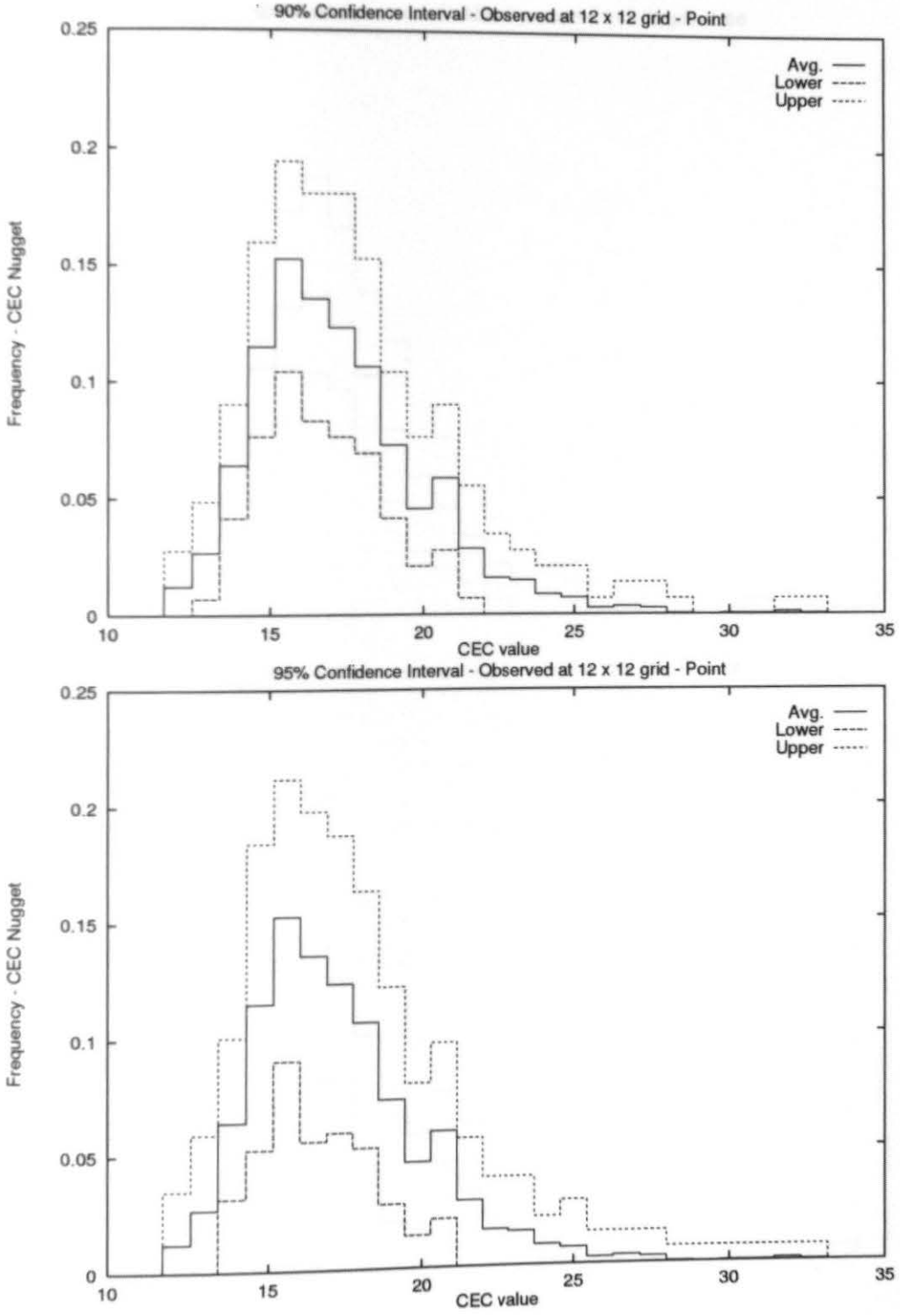


Figure C.37: Histogram Experiment 1 Nugget, 12x12 - Obs. point

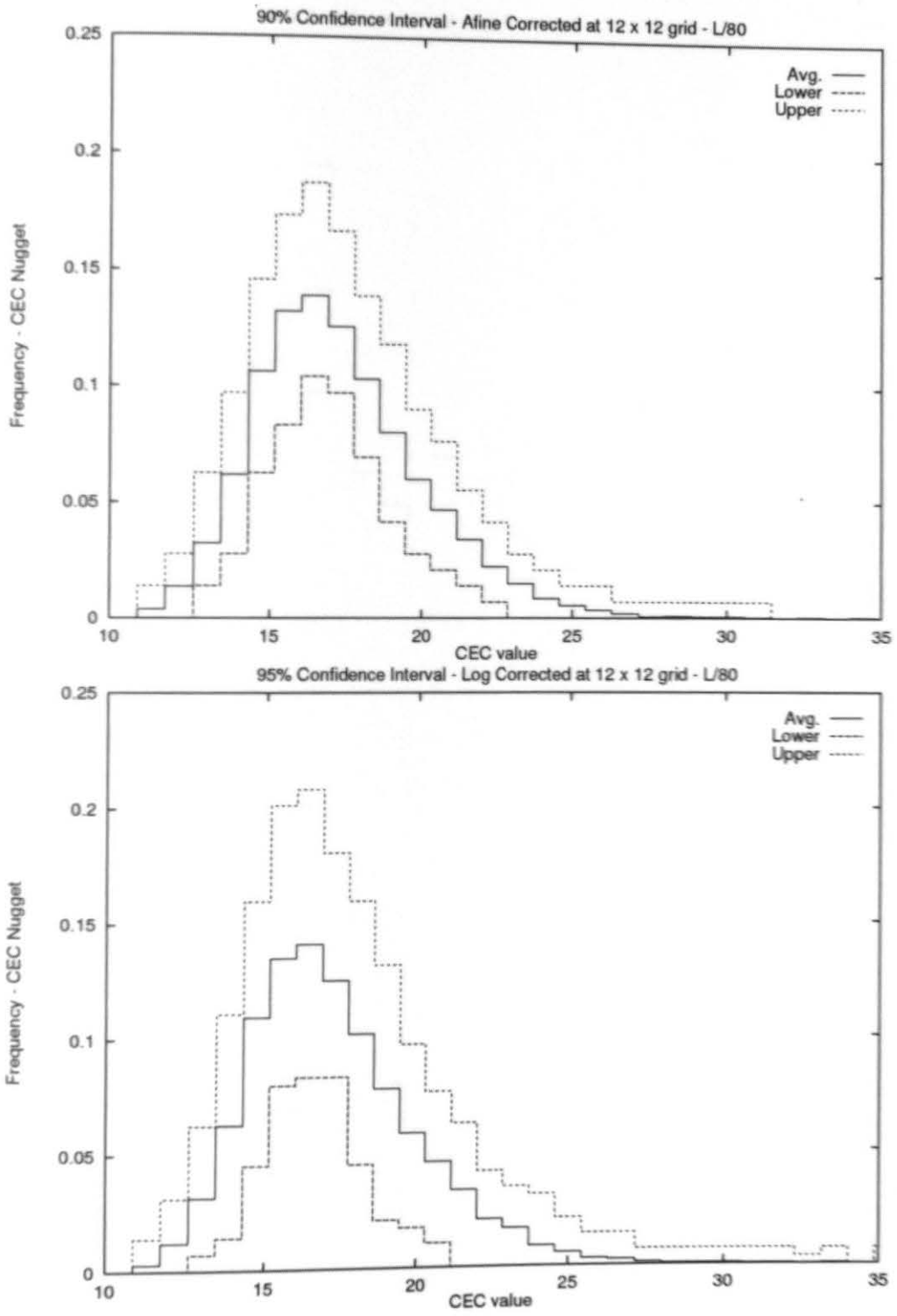


Figure C.38: Histogram Experiment 1 Nugget, 12x12 - Afine L/80

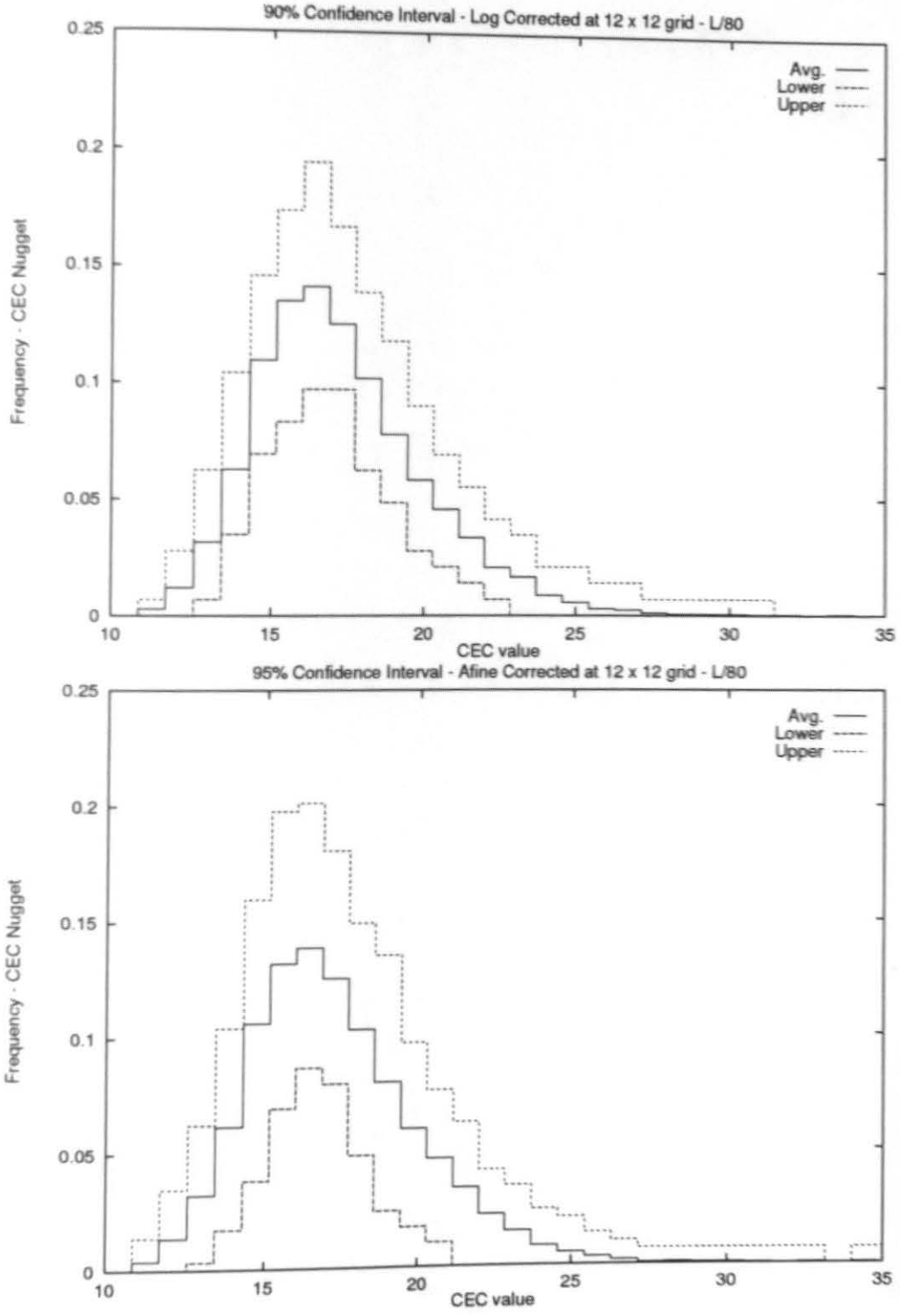


Figure C.39: Histogram Experiment 1 Nugget, 12x12 - Log L/80

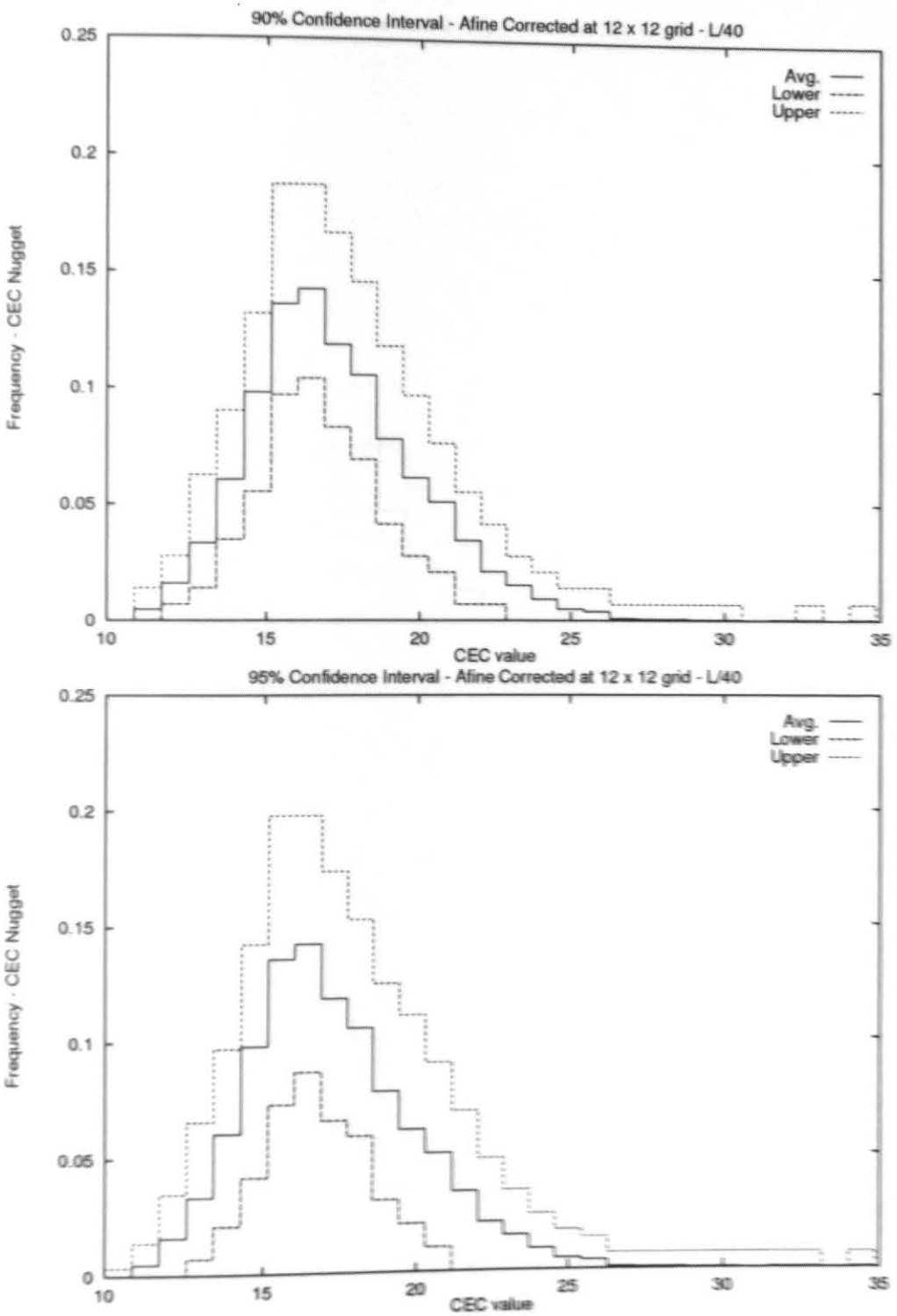


Figure C.40: Histogram Experiment 1 Nugget, 12x12 - Afine L/40

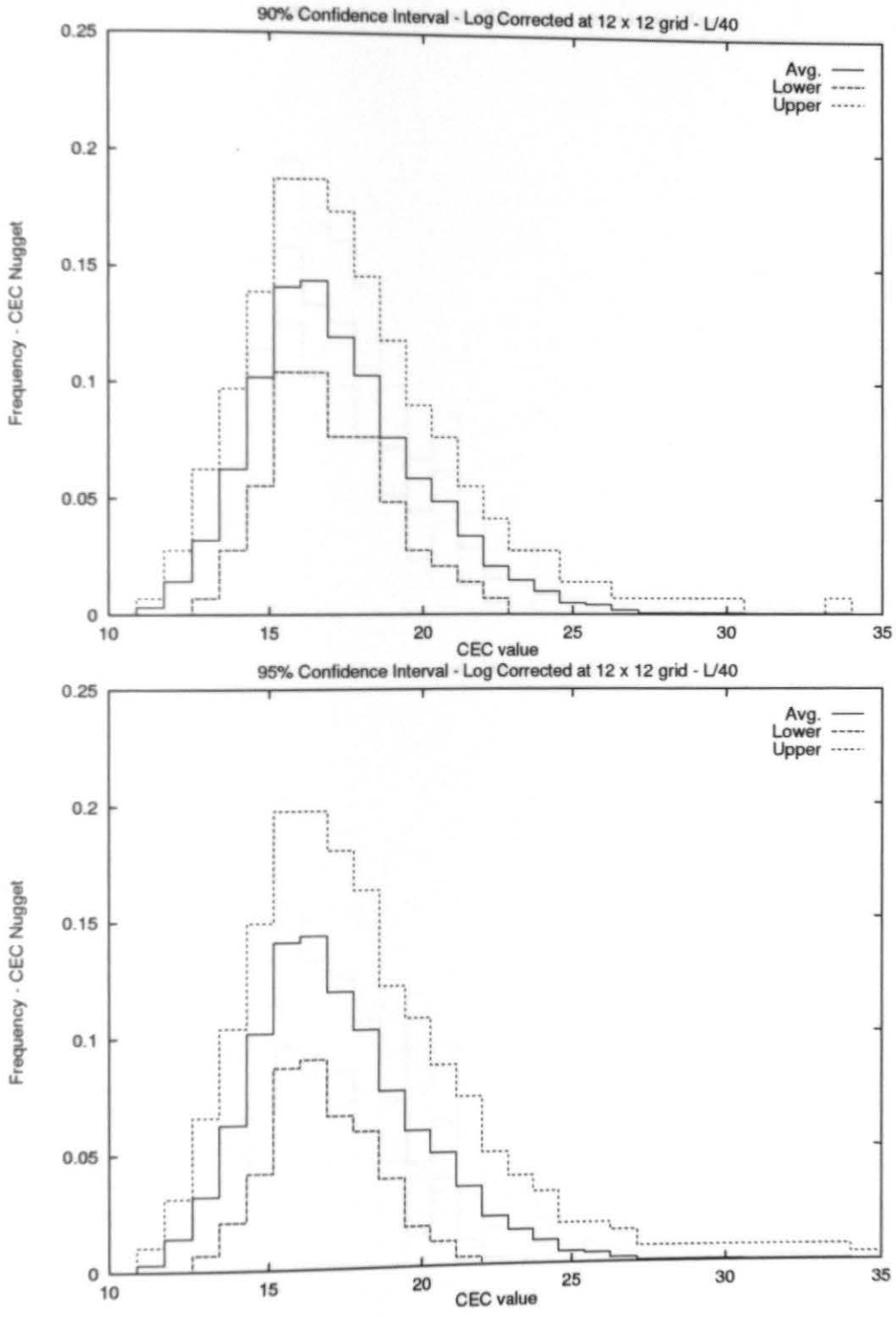


Figure C.41: Histogram Experiment 1 Nugget, 12x12 - Log L/40

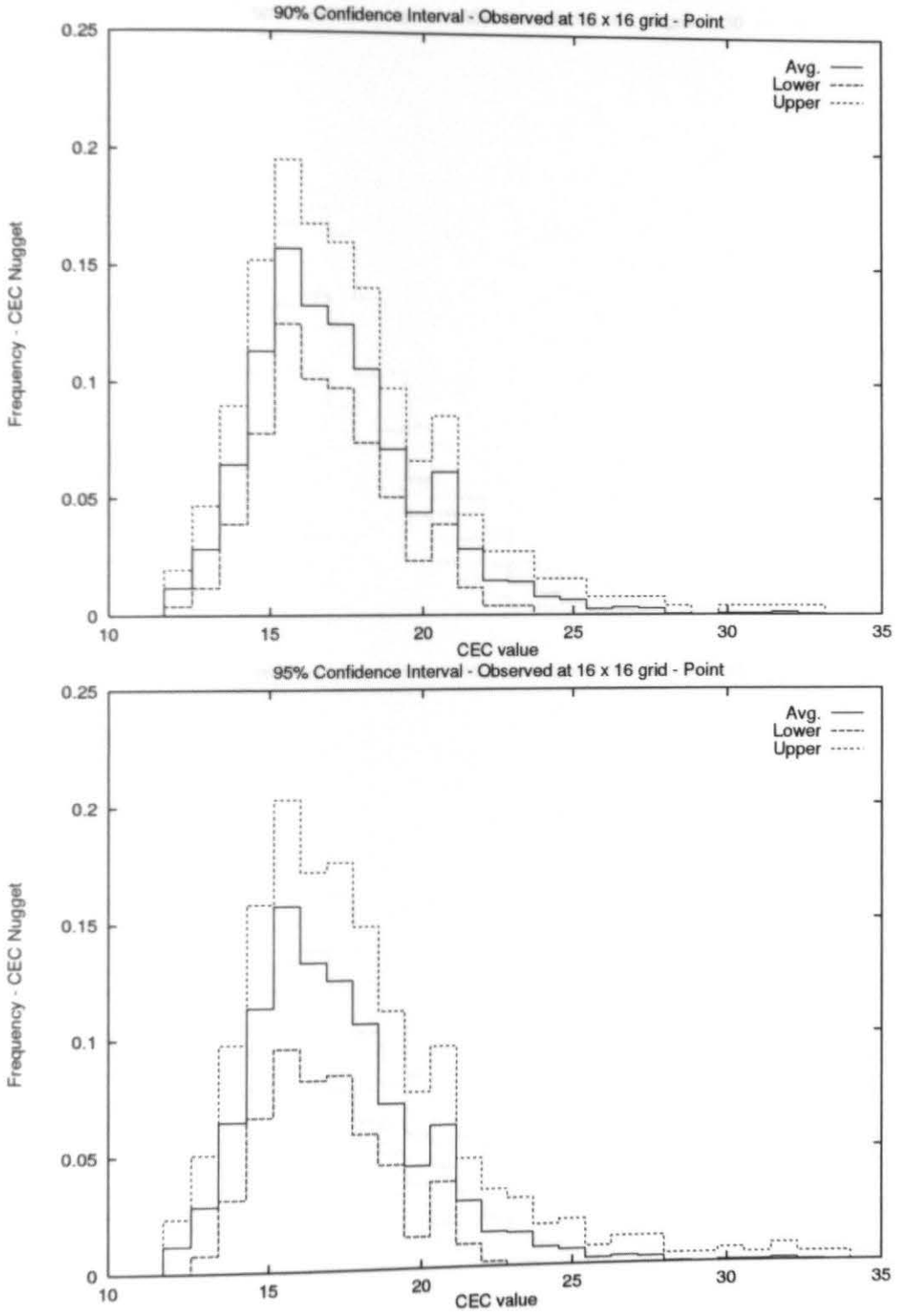


Figure C.42: Histogram Experiment 1 Nugget, 16x16 - Obs. point

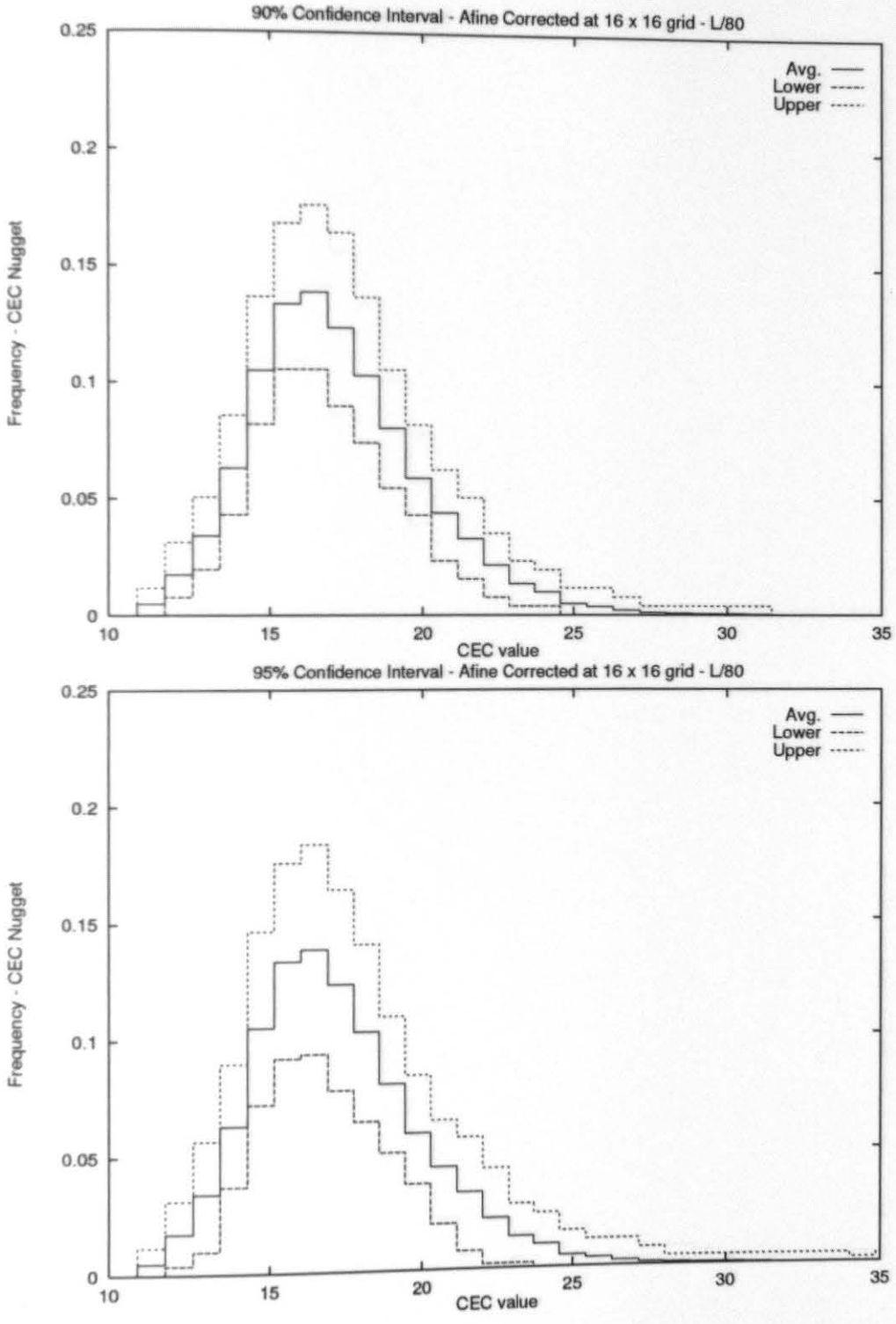


Figure C.43: Histogram Experiment 1 Nugget, 16x16 - Afine L/80

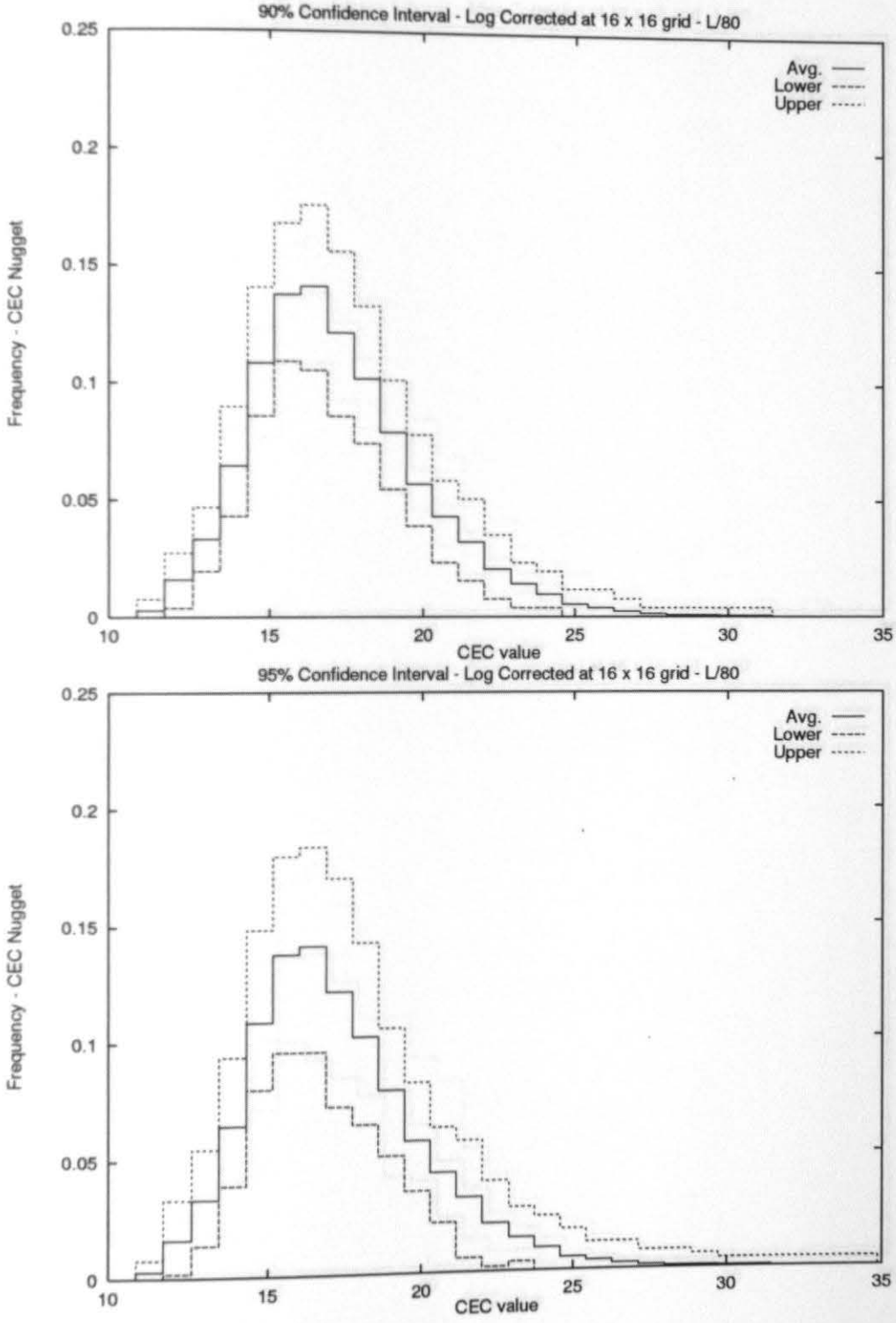


Figure C.44: Histogram Experiment 1 Nugget, 16x16 - Log L/80

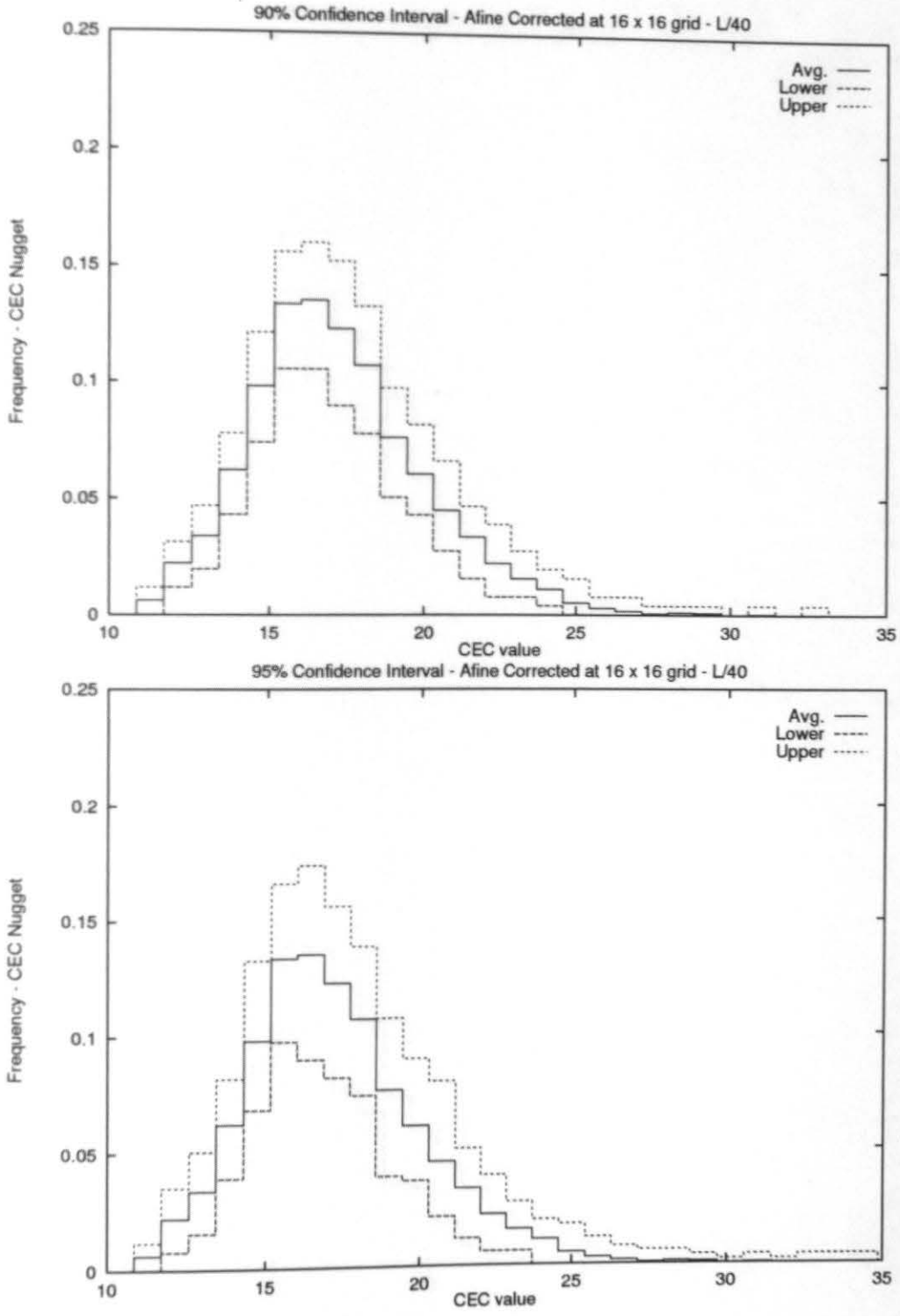


Figure C.45: Histogram Experiment 1 Nugget, 16x16 - Afine L/40

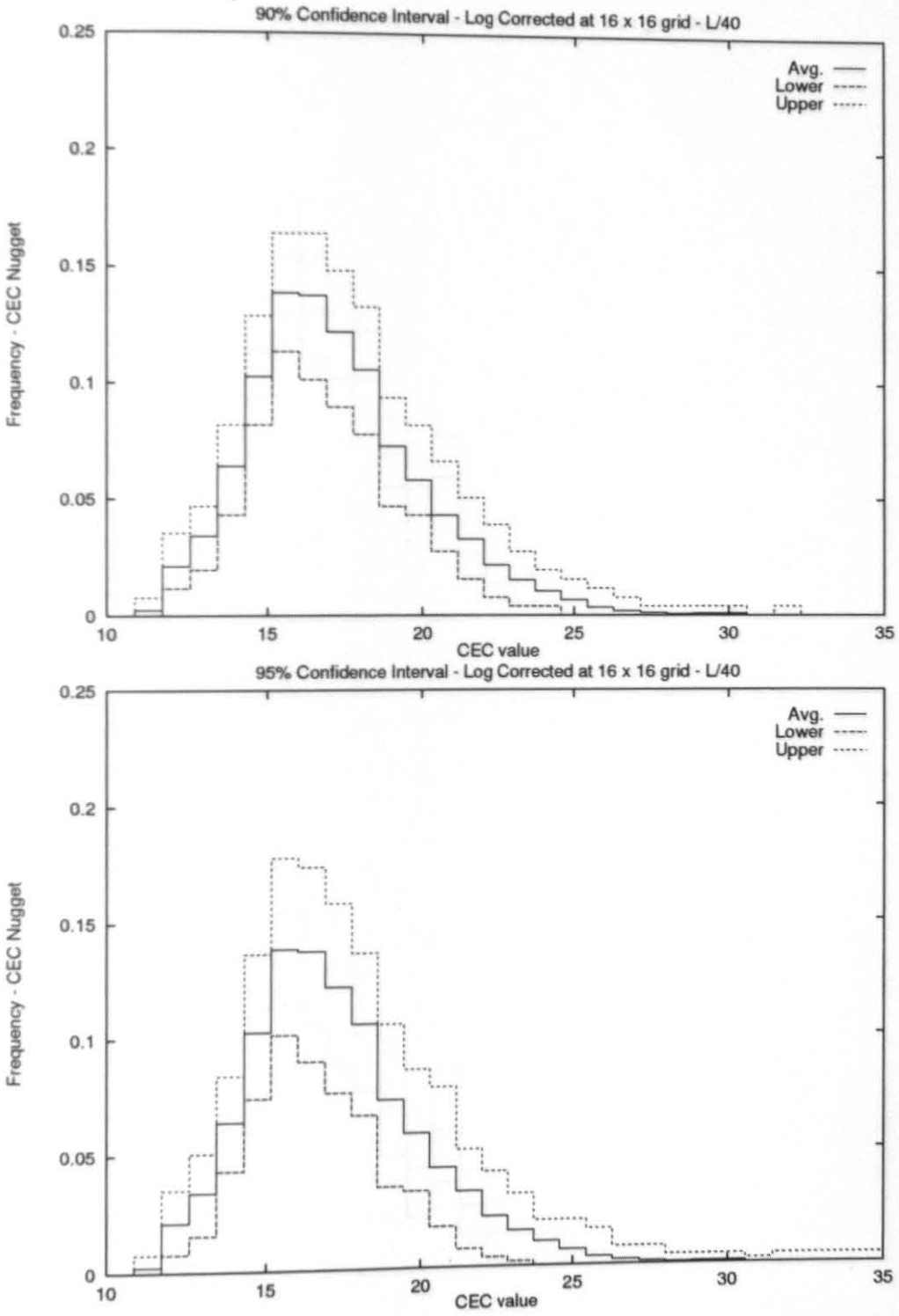


Figure C.46: Histogram Experiment 1 Nugget, 16x16 - Log L/40

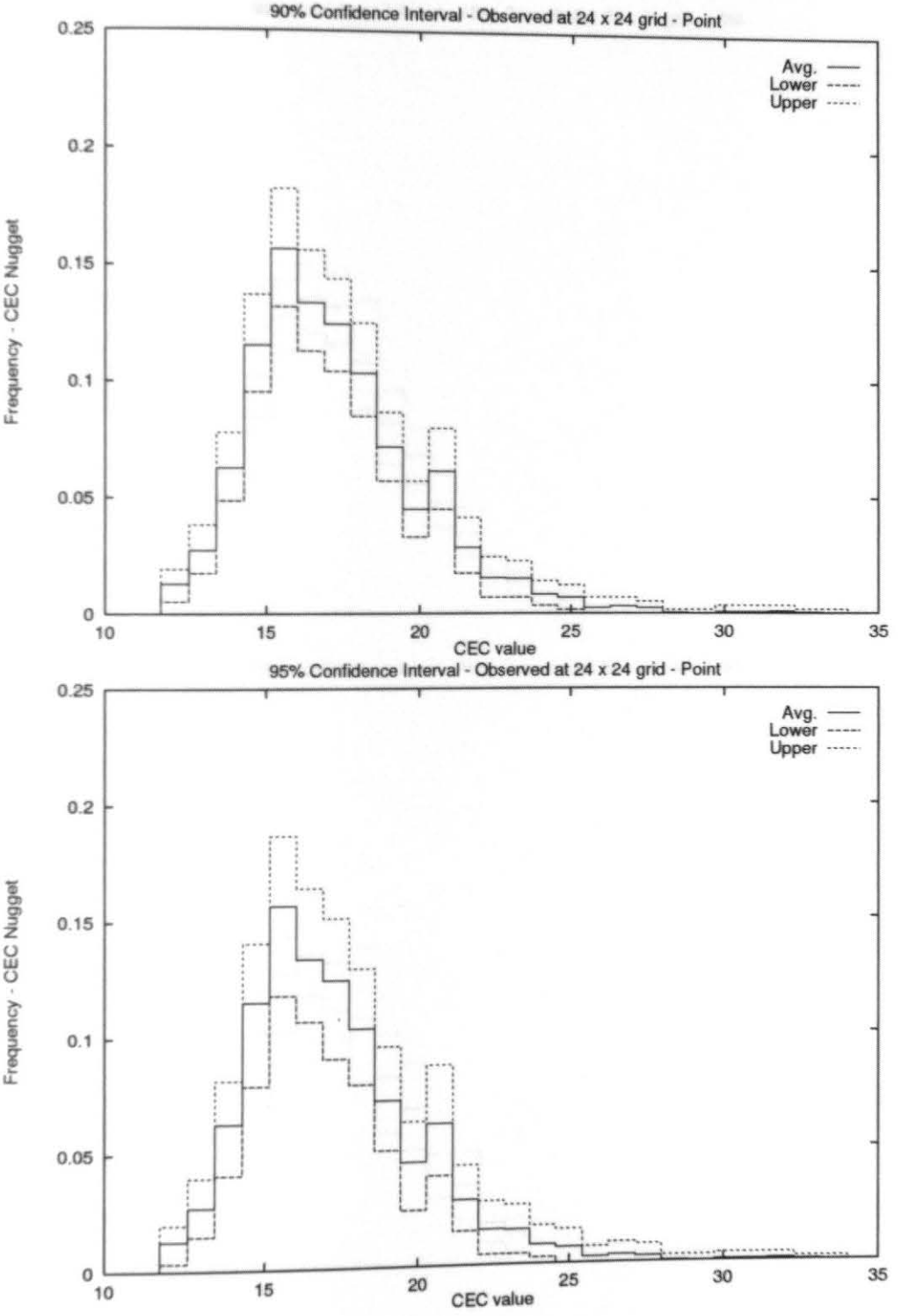


Figure C.47: Histogram Experiment 1 Nugget, 24x24 - Obs. point

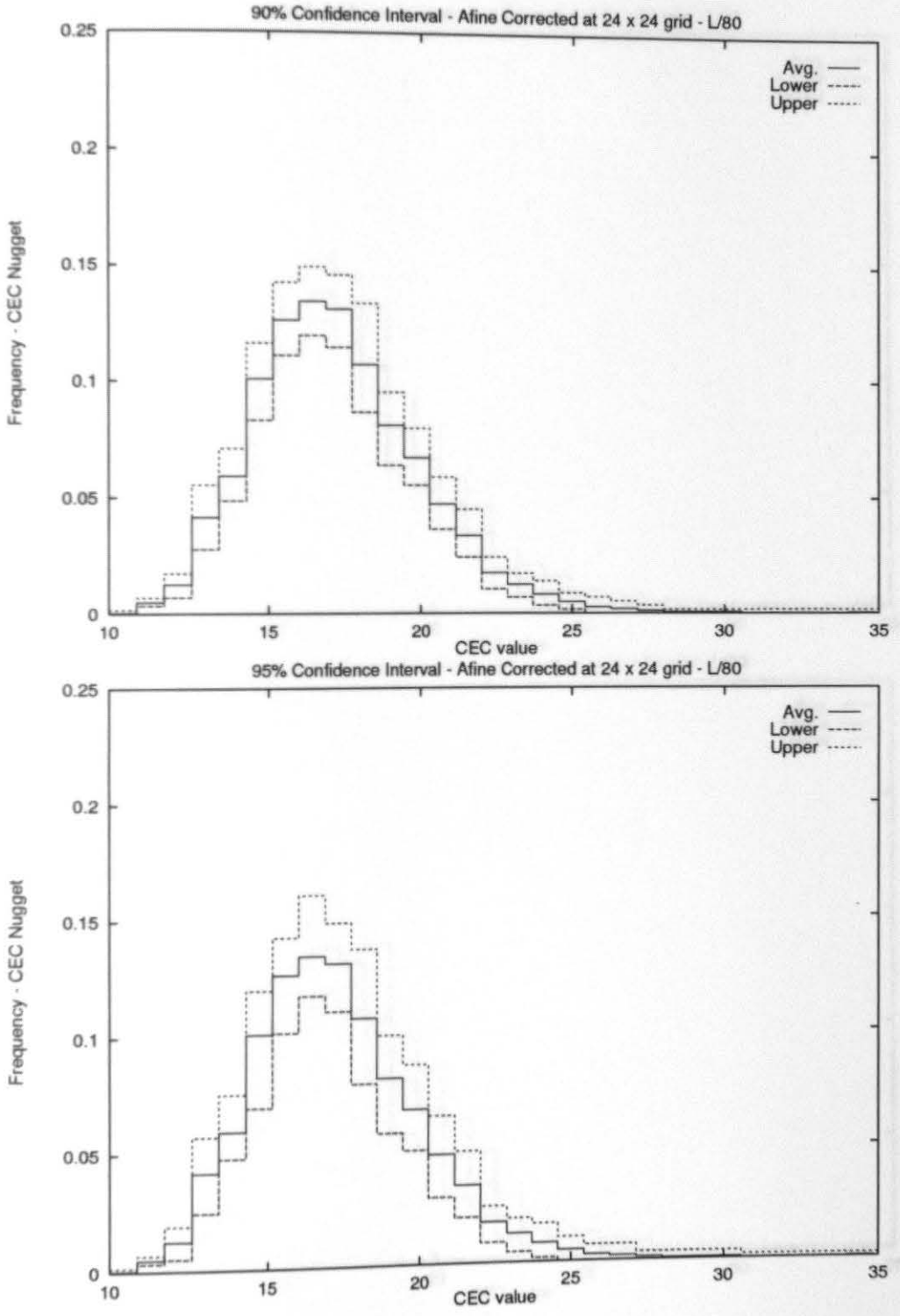


Figure C.48: Histogram Experiment 1 Nugget, 24x24 - Afine L/80

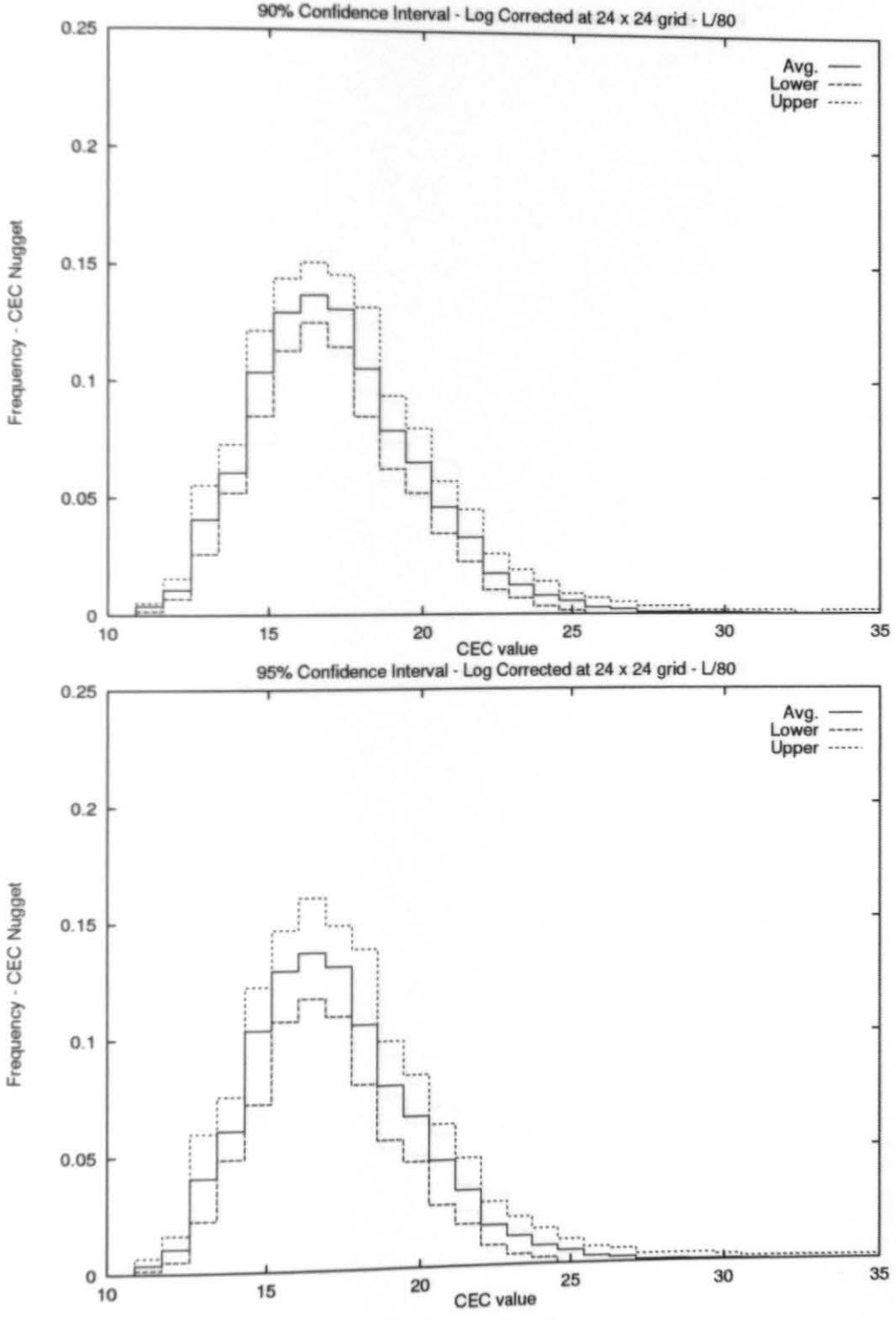


Figure C.49: Histogram Experiment 1 Nugget, 24x24 - Log L/80

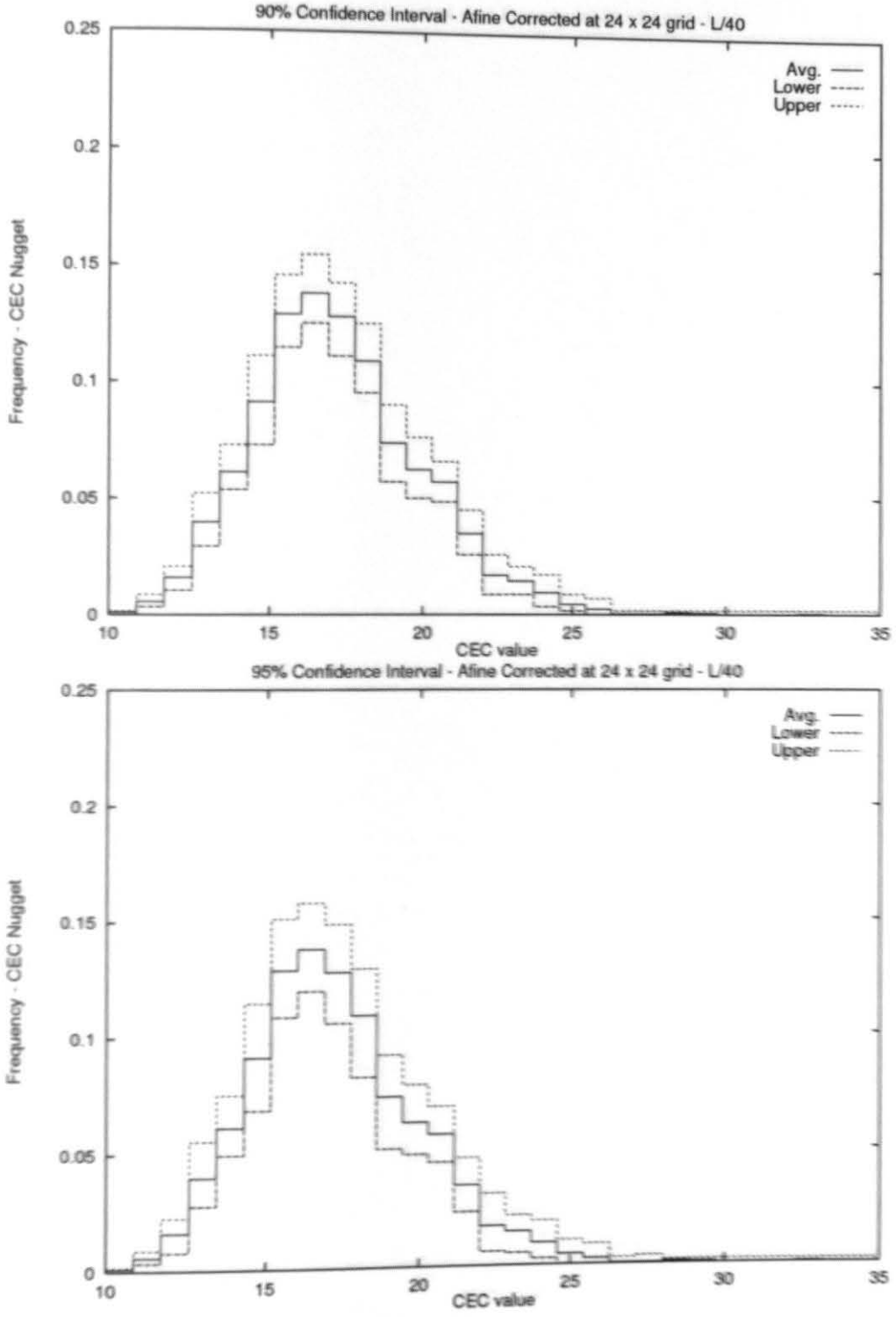


Figure C.50: Histogram Experiment 1 Nugget, 24x24 - Afine L/40

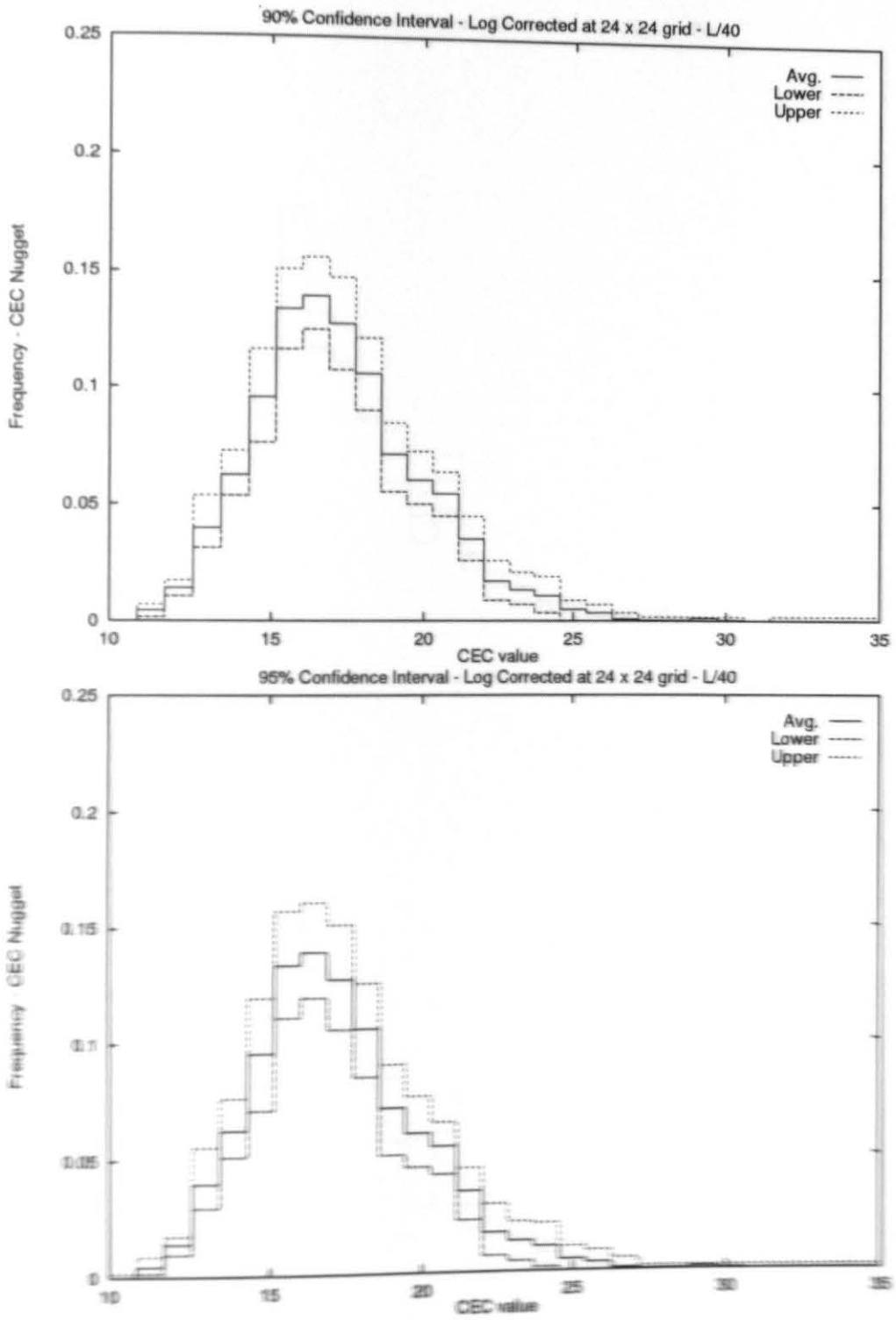


Figure C.51: Histogram Experiment 1 Nugget, 24x24 - Log L/40

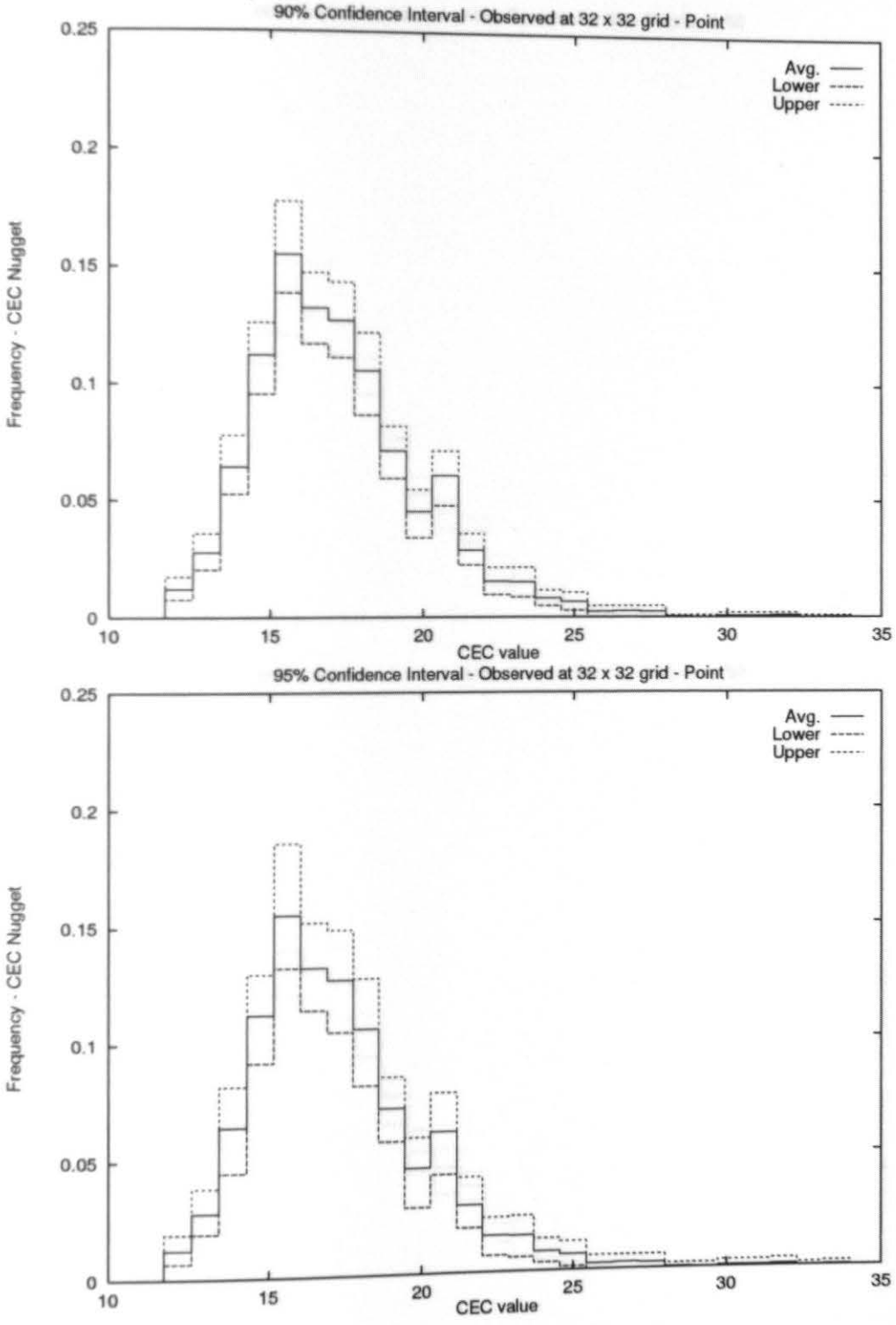


Figure C.52: Histogram Experiment 1 Nugget, 32x32 - Obs. point

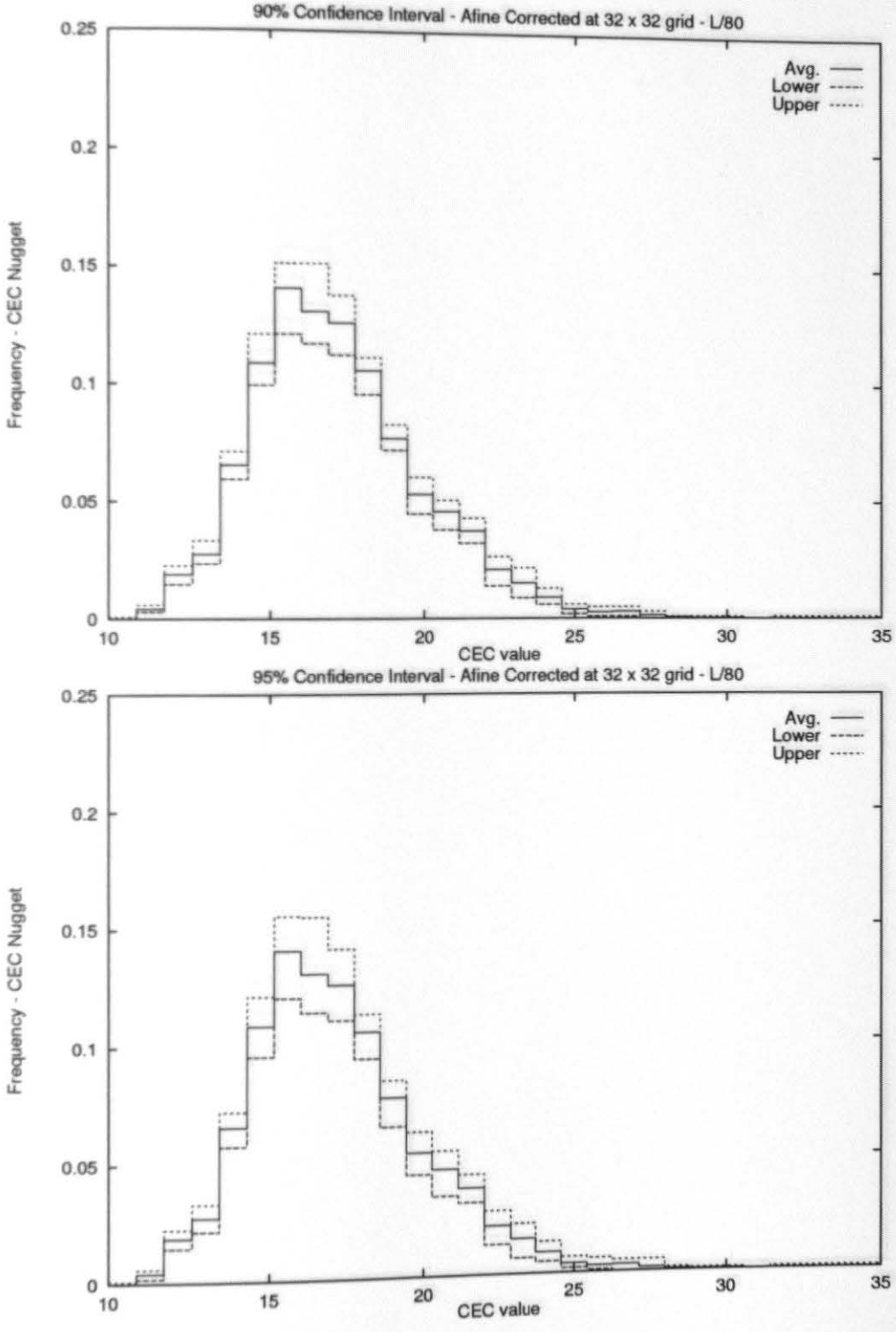


Figure C.53: Histogram Experiment 1 Nugget, 32x32 - Afine L/80

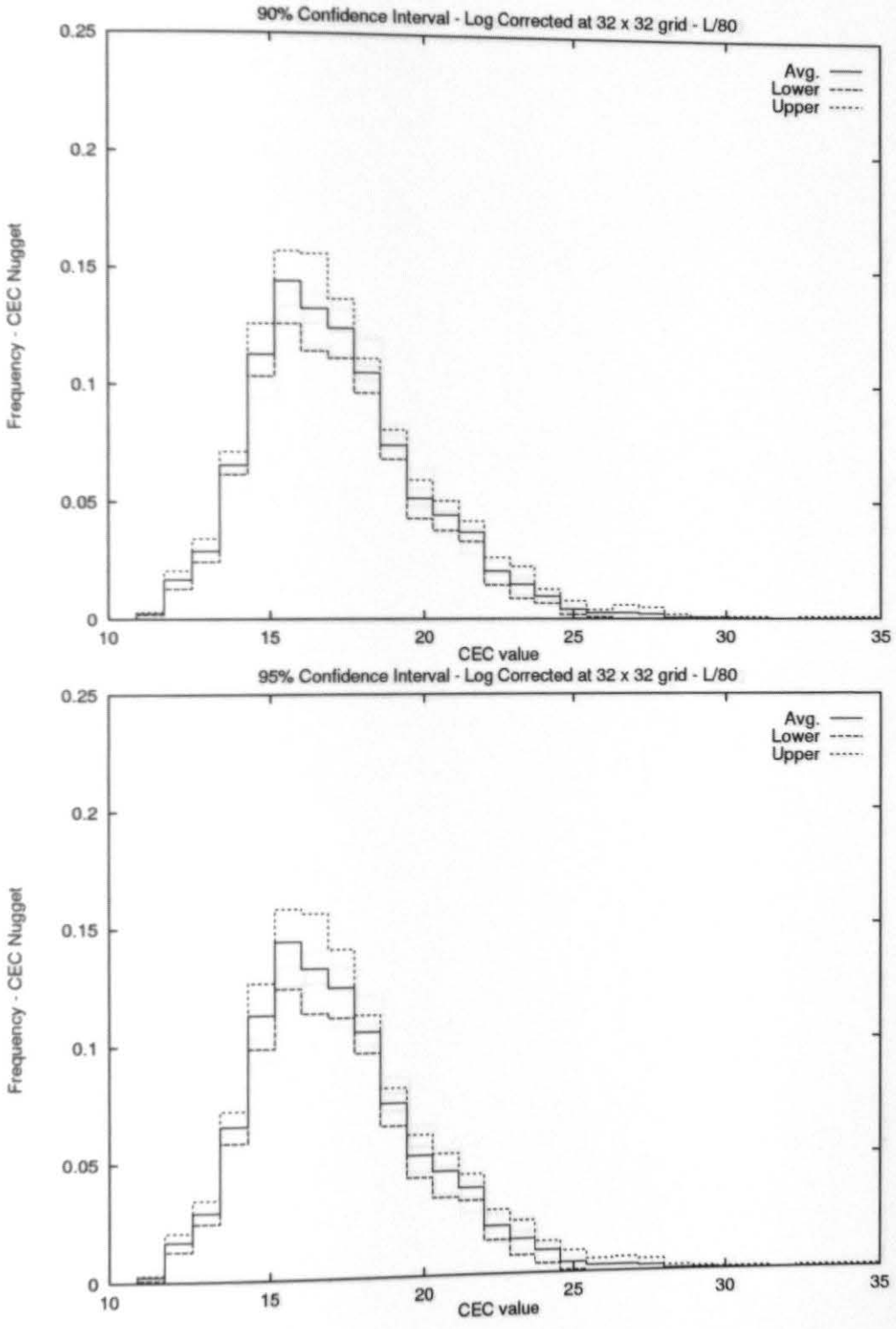


Figure C.54: Histogram Experiment 1 Nugget, 32x32 - Log L/80

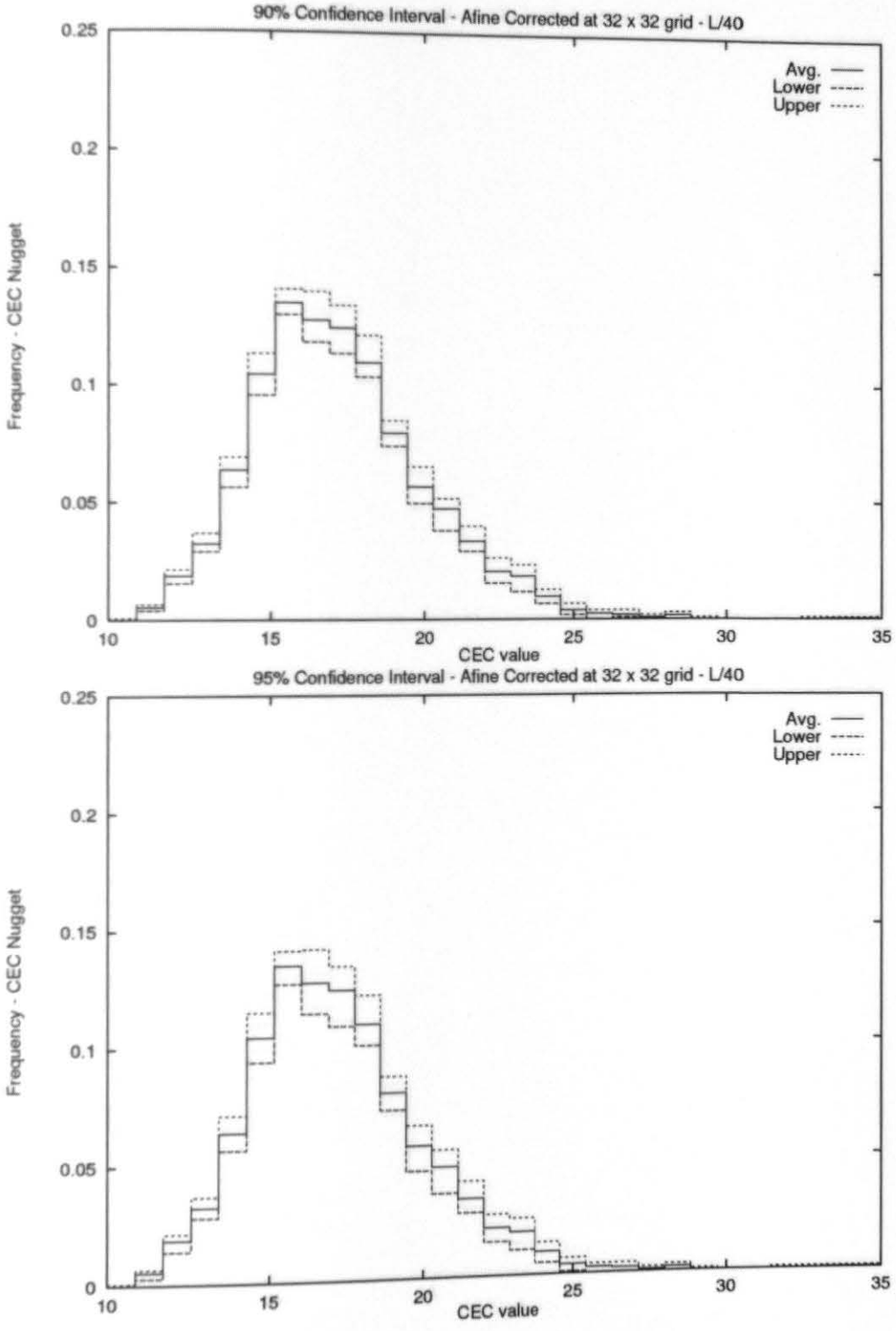


Figure C.55: Histogram Experiment 1 Nugget, 32x32 - Afine L/40

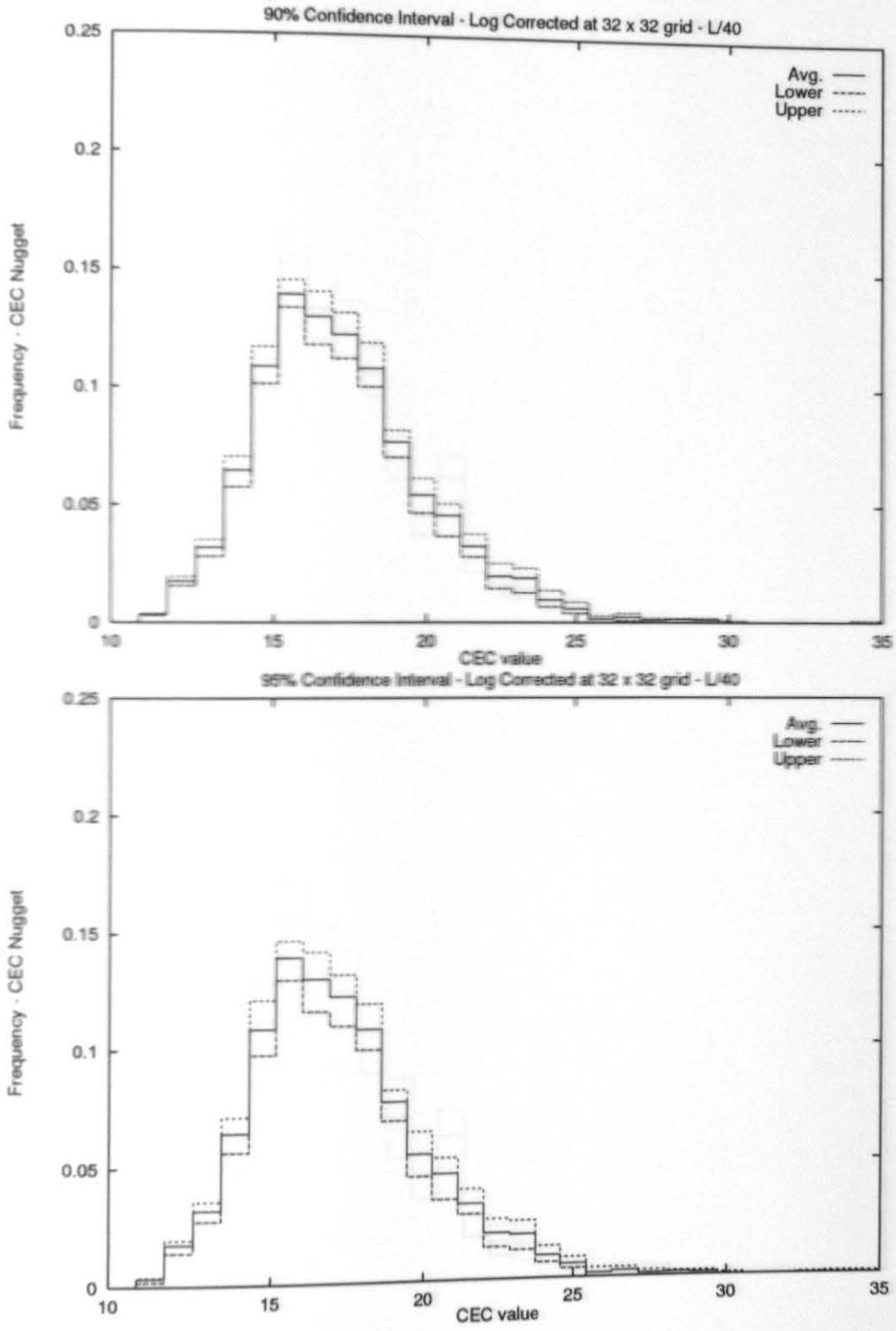


Figure C.56: Histogram Experiment 1 Nugget, 32x32 - Log L/40

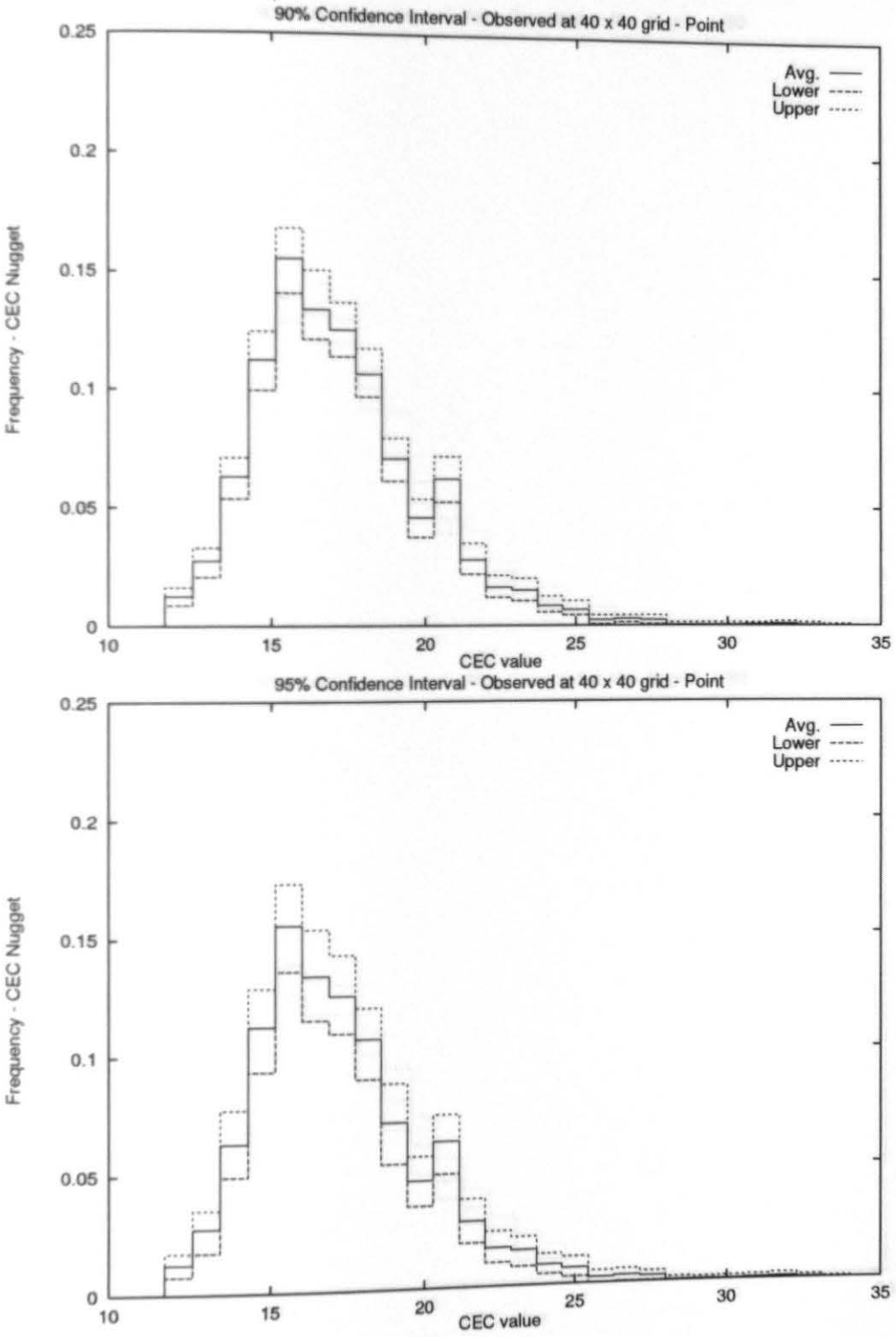


Figure C.57: Histogram Experiment 1 Nugget, 40x40 - Obs. point

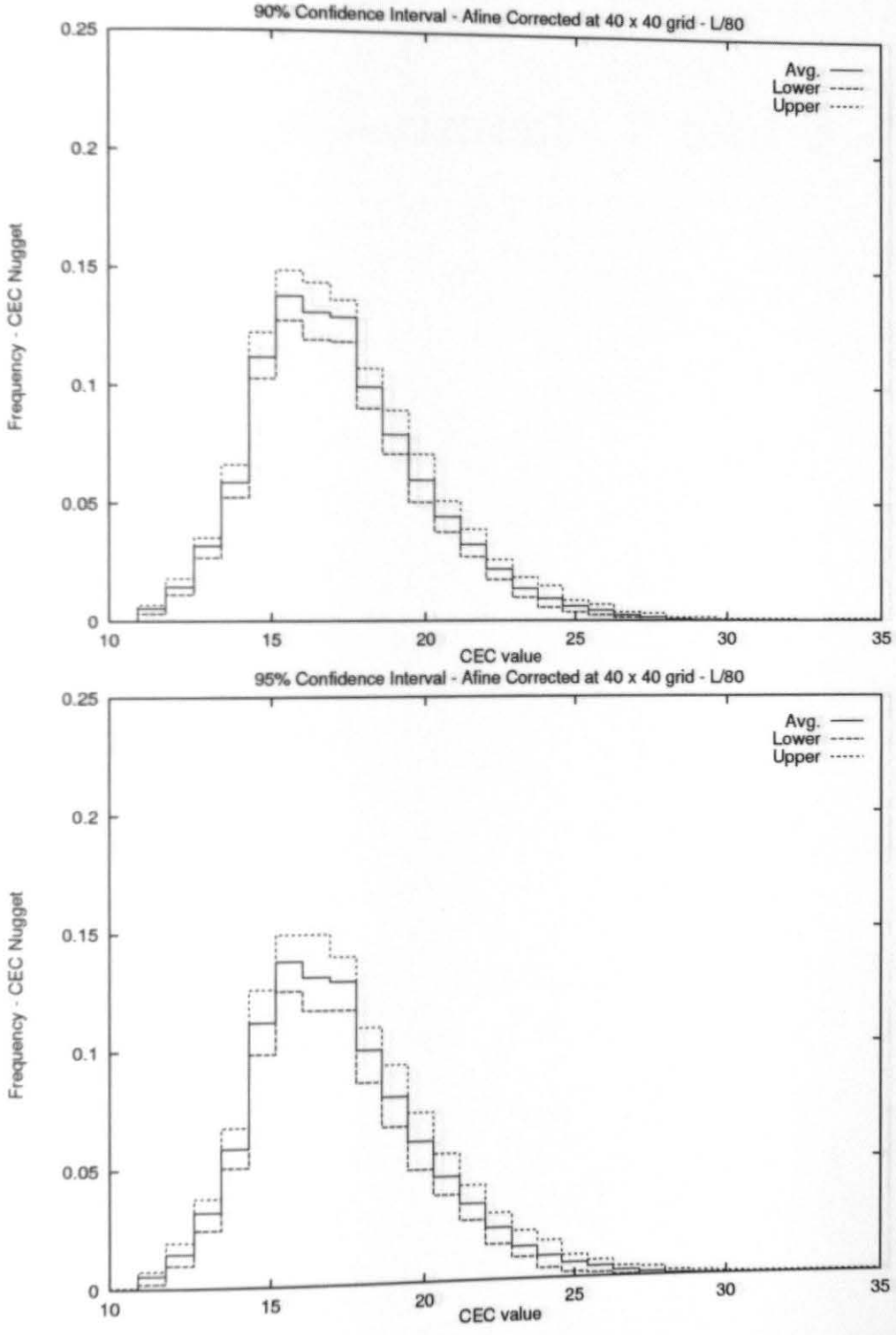


Figure C.58: Histogram Experiment 1 Nugget, 40x40 - Afine L/80

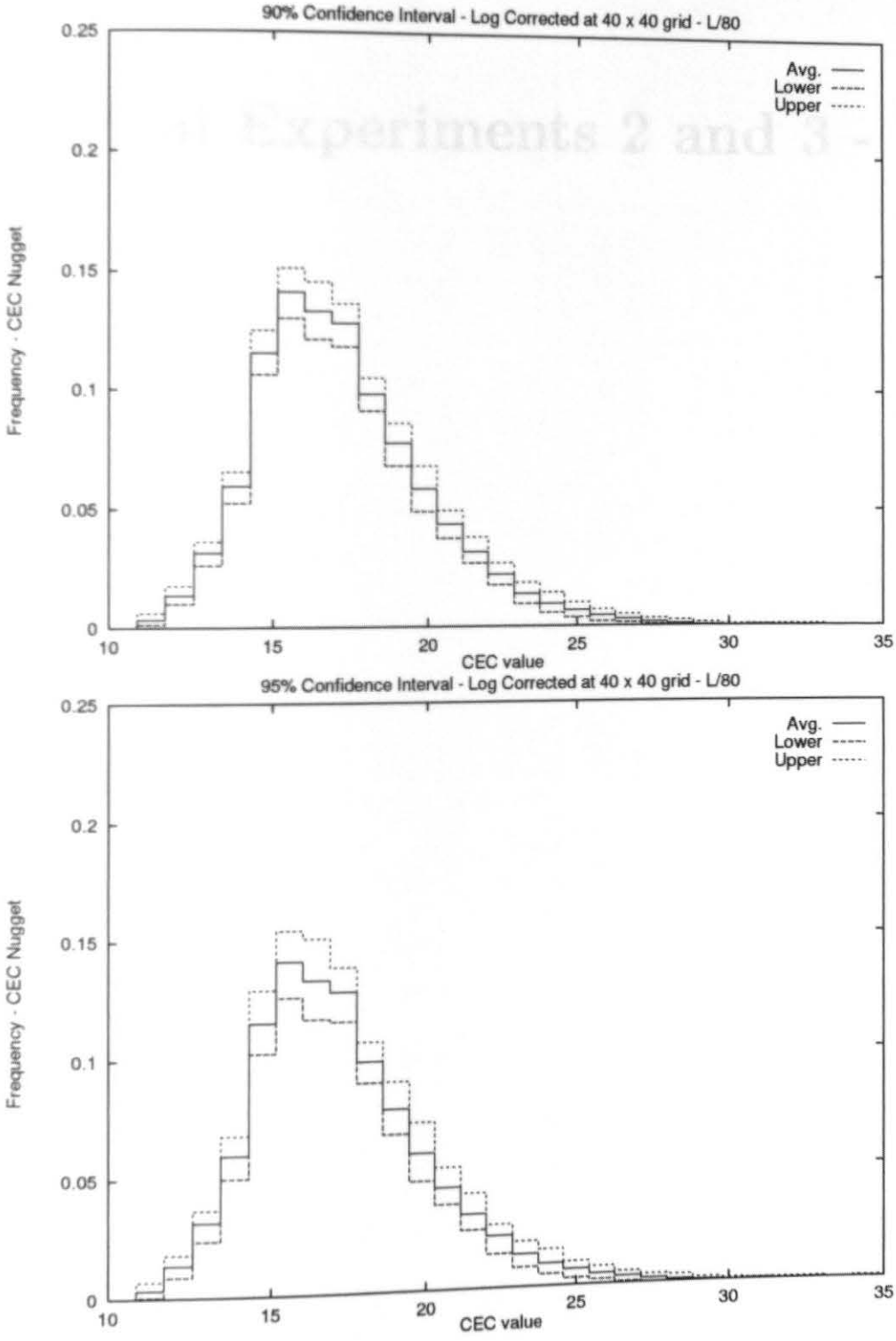


Figure C.59: Histogram Experiment 1 Nugget, 40x40 - Log L/80

Numerical Experiments 2 and 3 - Plain

Parameter	Mean	Bias	%Bias	S.D.	%S.D.	C.V.	Low C.I.	High C.I.	S.D. C.I.	Low C.I.	High C.I.
Point:											
Sill	7.144	0.079	1.109	1.213	16.977	0.170	6.986	7.302	0.497	5.086	9.931
Range	61.179	5.458	8.921	21.881	35.765	0.358	-498.052	620.409	1262.898	27.519	130.607
L/80:											
Sill	7.320	0.097	1.319	1.204	16.449	0.164	6.630	8.009	7.091	5.369	10.331
Range	58.187	2.466	4.239	16.285	27.987	0.280	-295.033	411.408	933.784	26.862	92.857
L/40:											
Sill	7.360	0.137	1.866	1.494	20.295	0.203	4.465	10.256	47.758	5.405	10.241
Range	62.626	6.905	11.026	16.415	26.210	0.262	-245.952	371.205	705.086	33.148	102.639
Mixed: pt & L/80											
Sill	7.163	0.060	0.845	0.775	10.825	0.108	7.051	7.274	0.420	5.672	8.789
Range	60.154	4.433	7.370	18.406	30.598	0.306	-322.552	442.860	575.363	31.141	115.214
Mixed: pt & L/40											
Sill	7.158	0.065	0.906	0.647	9.041	0.090	7.025	7.291	0.407	5.923	8.677
Range	63.427	7.706	12.150	15.800	24.911	0.249	-357.861	484.715	655.019	35.810	96.960
Mixed: L/80 & L/40											
Sill	7.296	0.073	0.994	0.670	9.184	0.092	6.824	7.767	2.960	6.181	8.866
Range	61.099	5.378	8.802	16.388	26.822	0.268	-260.292	382.490	492.427	32.792	99.313
pt + L/80											
Sill	7.321	0.098	1.338	0.880	12.024	0.120	7.222	7.420	0.257	5.779	9.189
Range	57.995	2.274	3.921	26.625	45.910	0.459	-276.968	392.958	1293.871	16.215	114.224
pt + L/40											
Sill	7.365	0.142	1.925	0.771	10.469	0.105	7.270	7.459	0.226	5.994	9.210
Range	66.503	10.782	16.212	33.072	49.731	0.497	-289.967	422.972	1410.112	24.860	136.079
L/80 + L/40											
Sill	7.370	0.147	1.996	0.631	8.560	0.086	7.293	7.447	0.117	6.580	9.040
Range	63.642	7.921	12.446	27.075	42.543	0.425	-178.815	306.099	627.730	21.984	113.505
All variograms											
Sill	7.334	0.111	1.519	0.581	7.923	0.079	7.303	7.366	0.044	6.599	8.856
Range	60.999	5.278	8.653	22.990	37.689	0.377	-24.129	146.127	176.087	22.414	99.176

Table D.1: Experiment 2 Plain, 24x24, Sill = 7.223, Range = 3*55.721. Confidence Intervals (C.I.) are at the 95% level

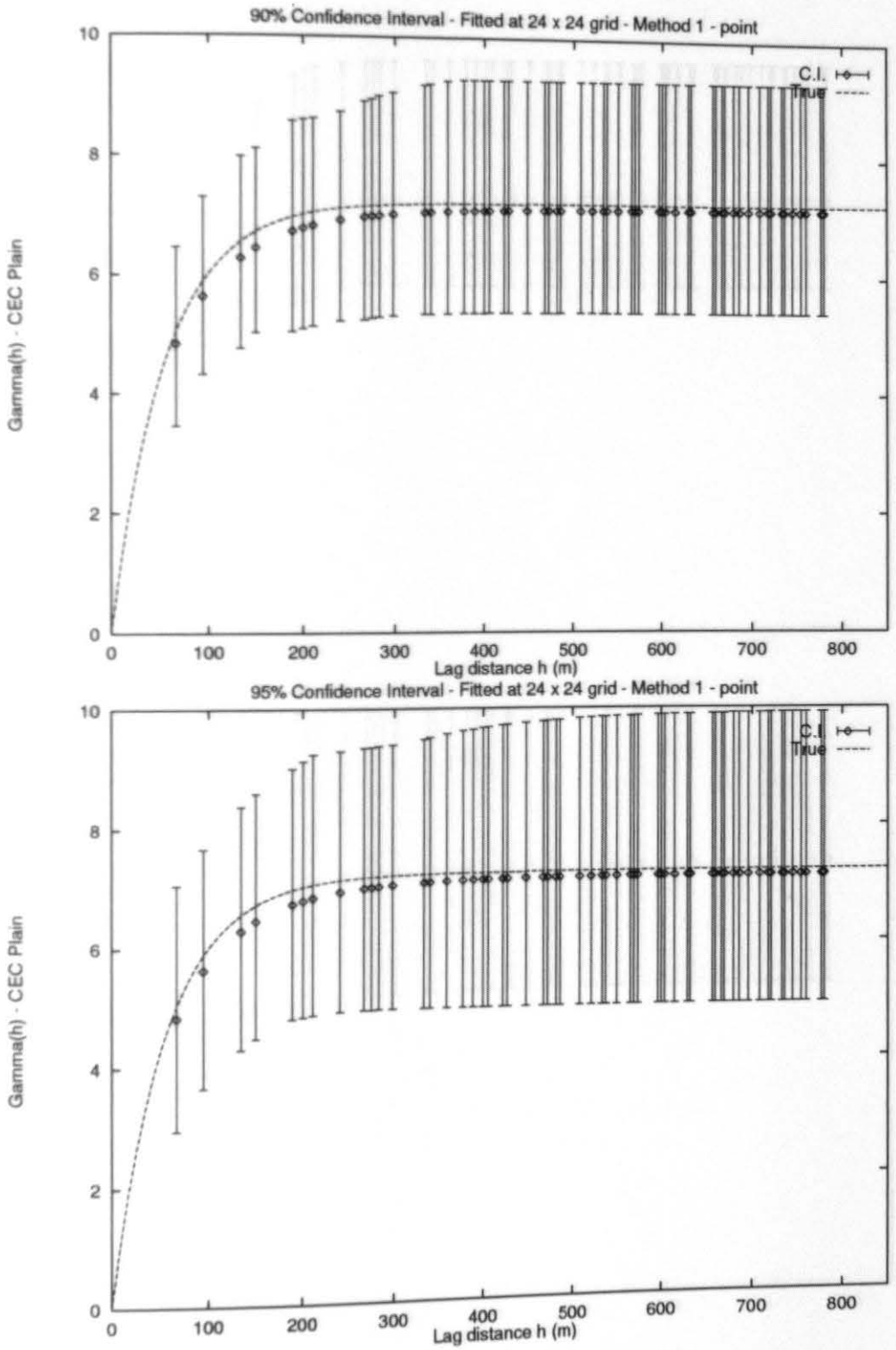


Figure D.1: Numerical Experiment 2 Plain, (192 points) - Point

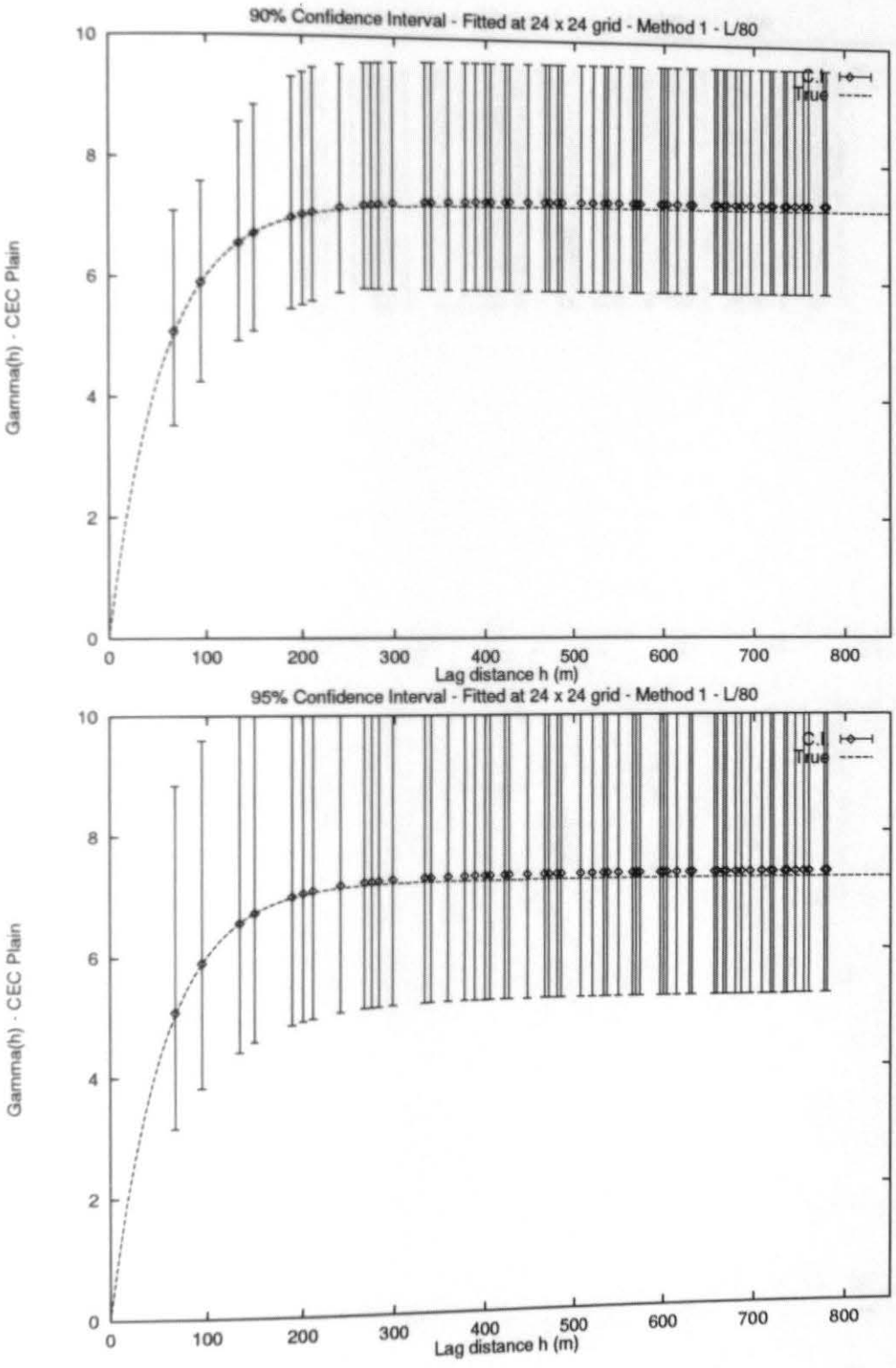


Figure D.2: Numerical Experiment 2 Plain, (192 points) - L/80

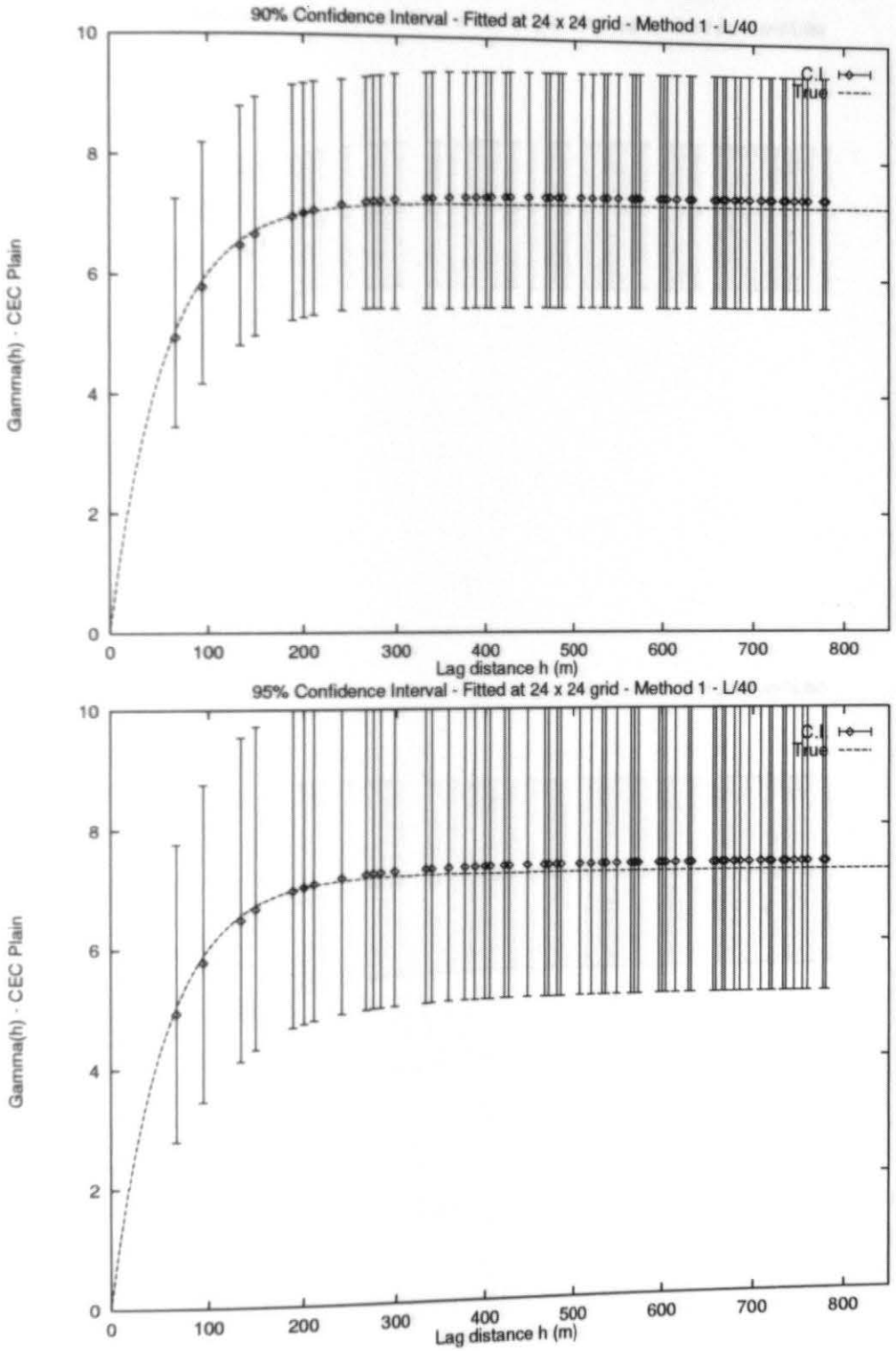


Figure D.3: Numerical Experiment 2 Plain, (192 points) - L/40

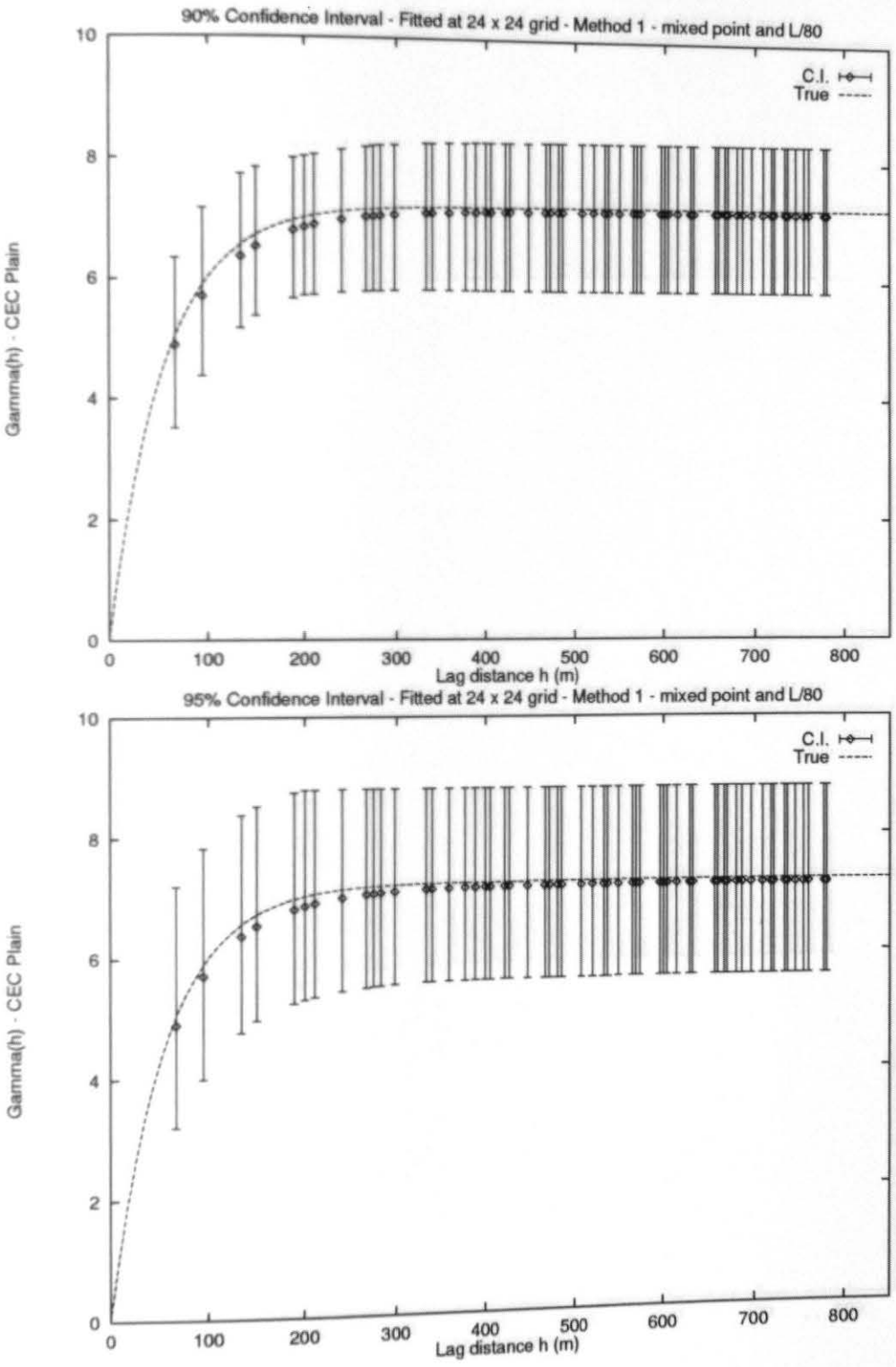


Figure D.4: Numerical Experiment 2 Plain, (192 points) - mixed point and L/80

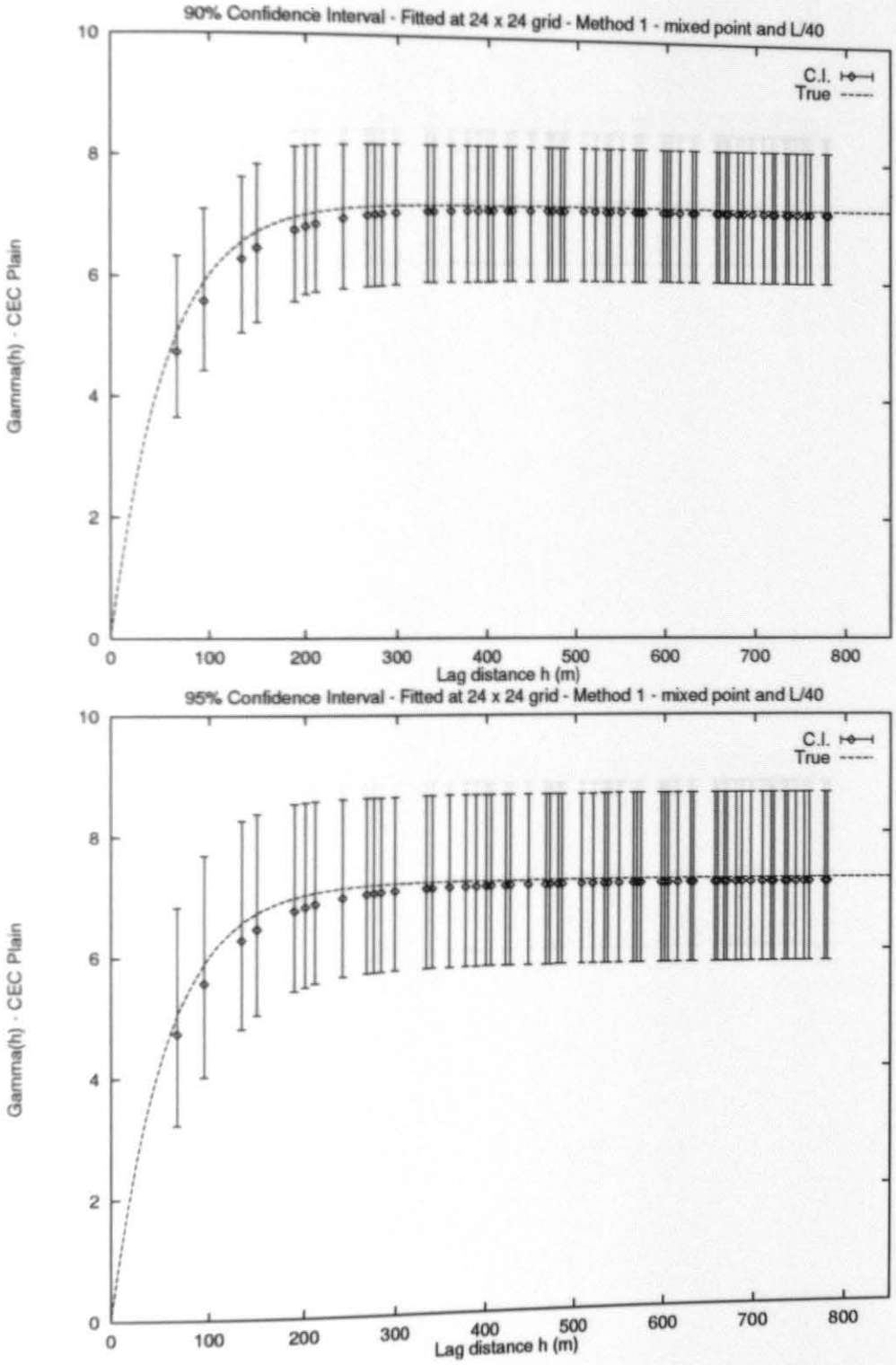


Figure D.5: Numerical Experiment 2 Plain, (192 points) - mixed point and L/40

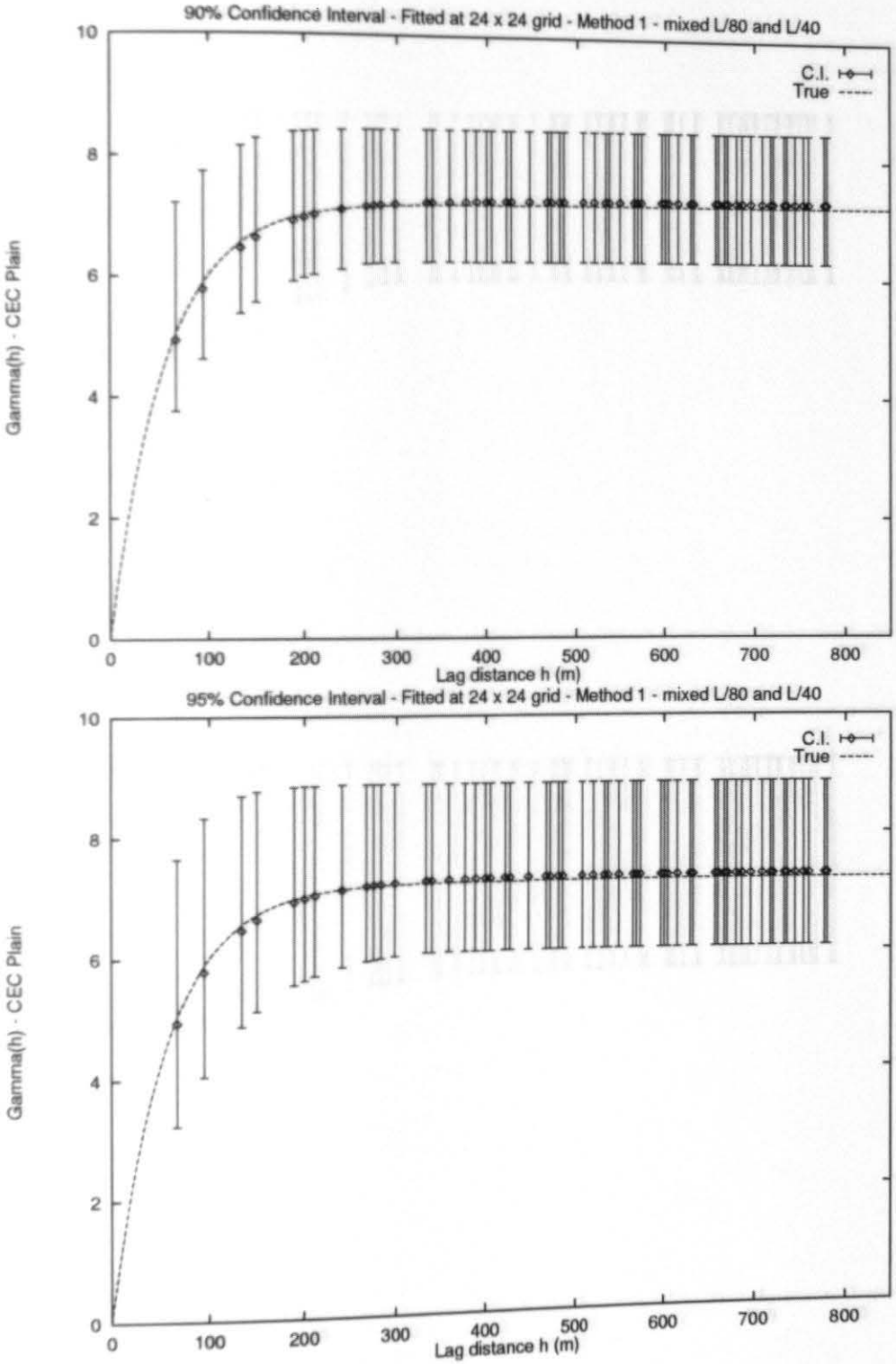


Figure D.6: Numerical Experiment 2 Plain, (192 points) - mixed L/80 and L/40

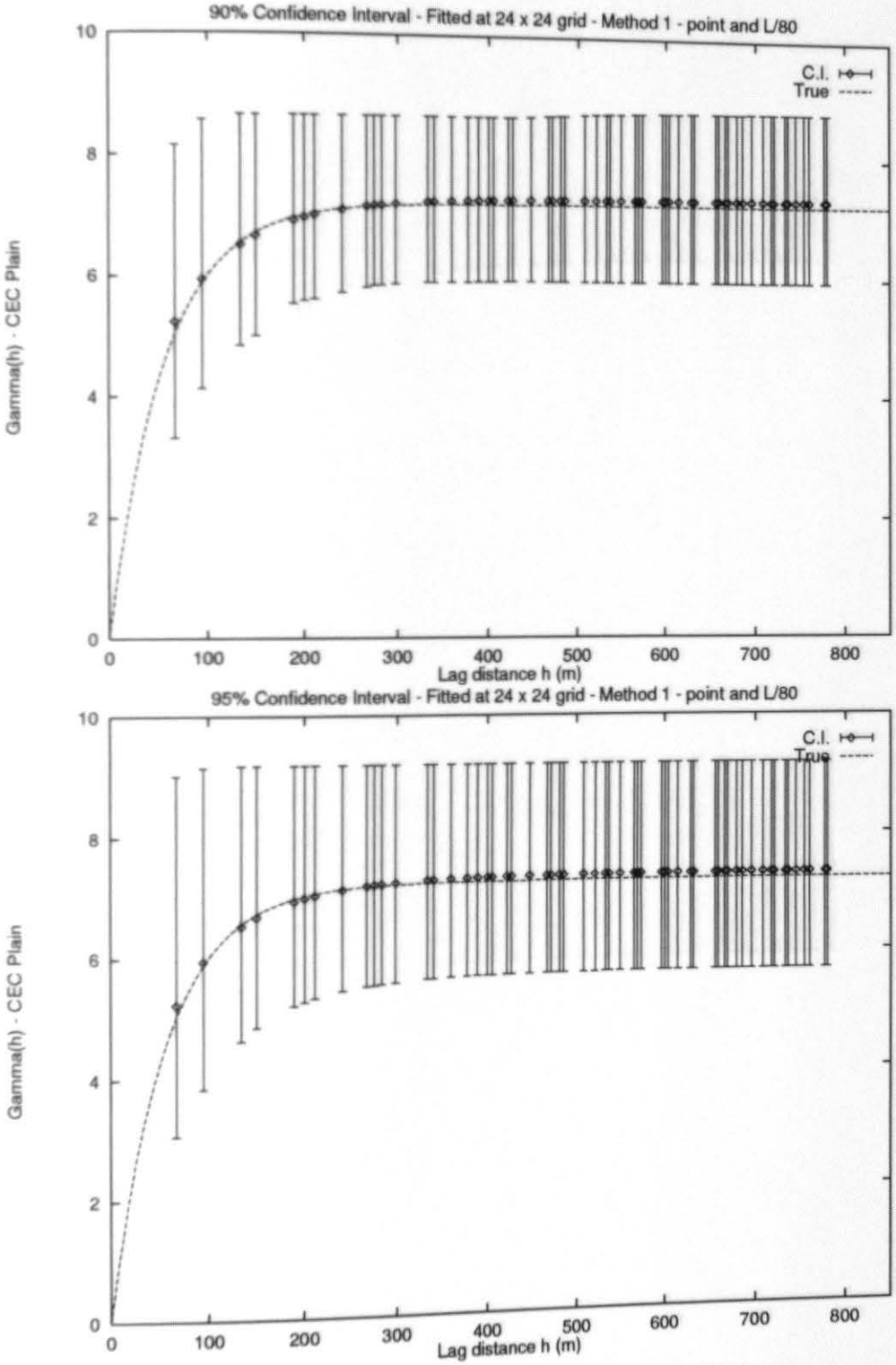


Figure D.7: Numerical Experiment 2 Plain, (192 points) - point and L/80

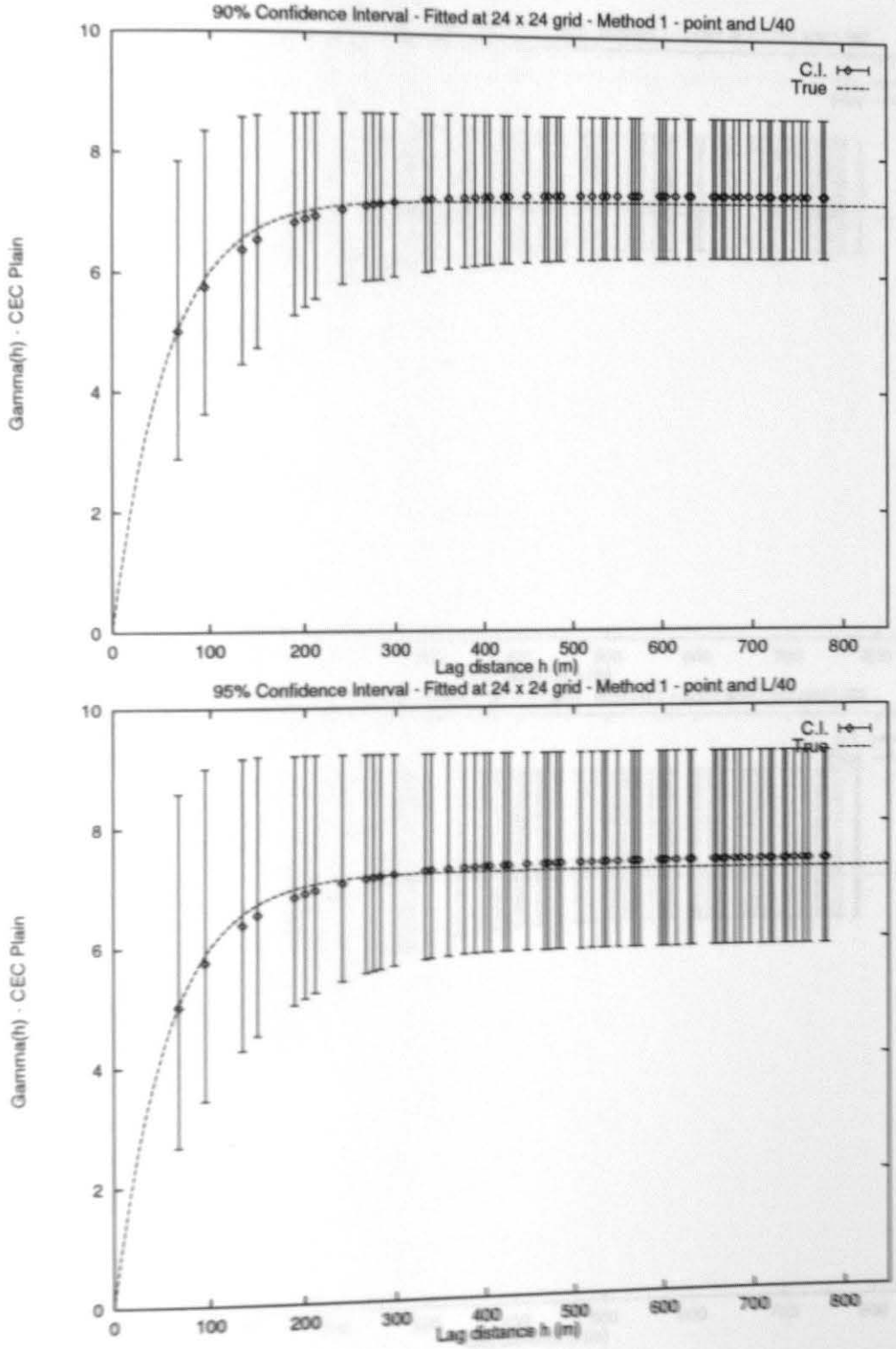


Figure D.8: Numerical Experiment 2 Plain, (192 points) - point and L/40

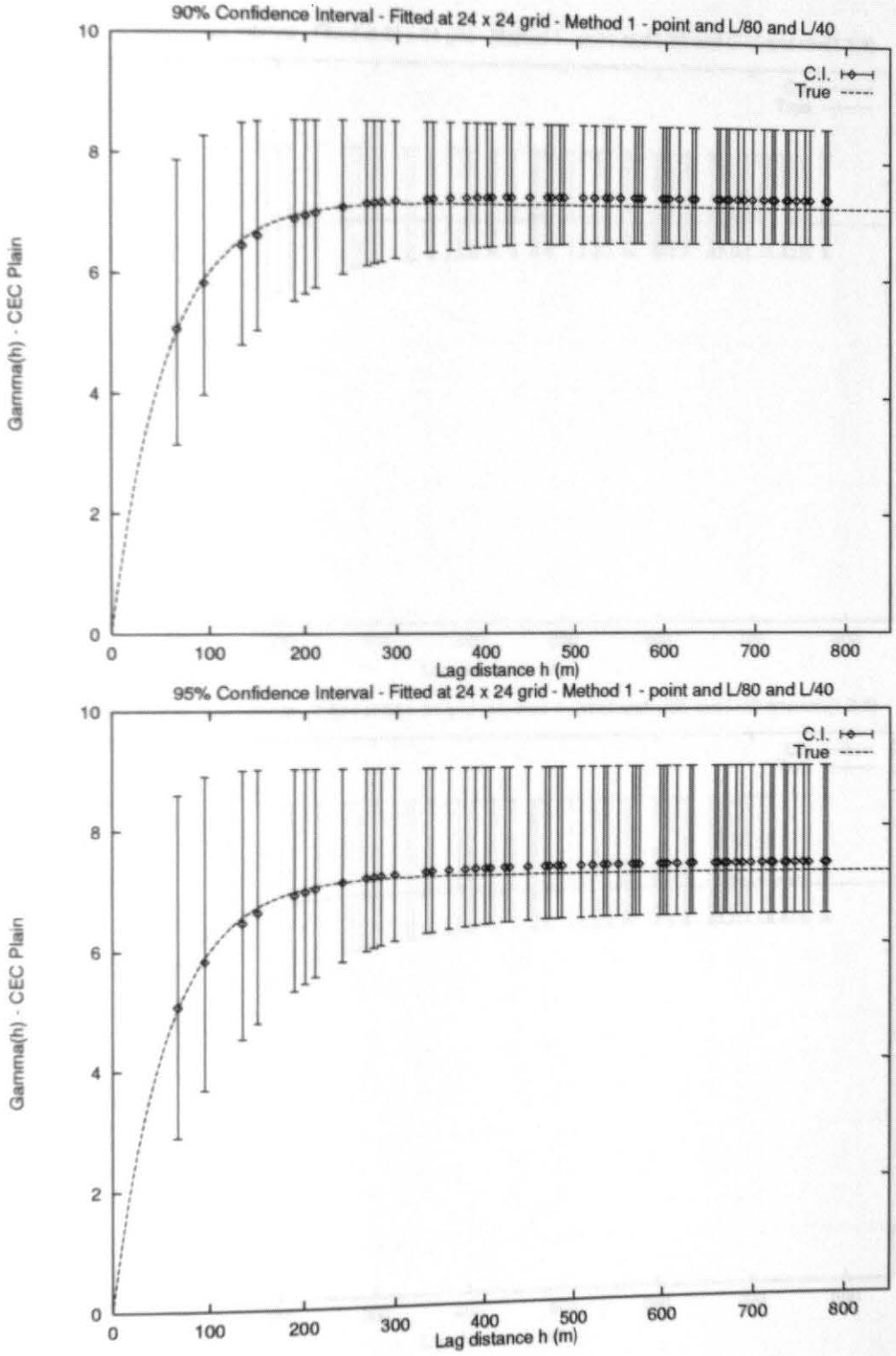
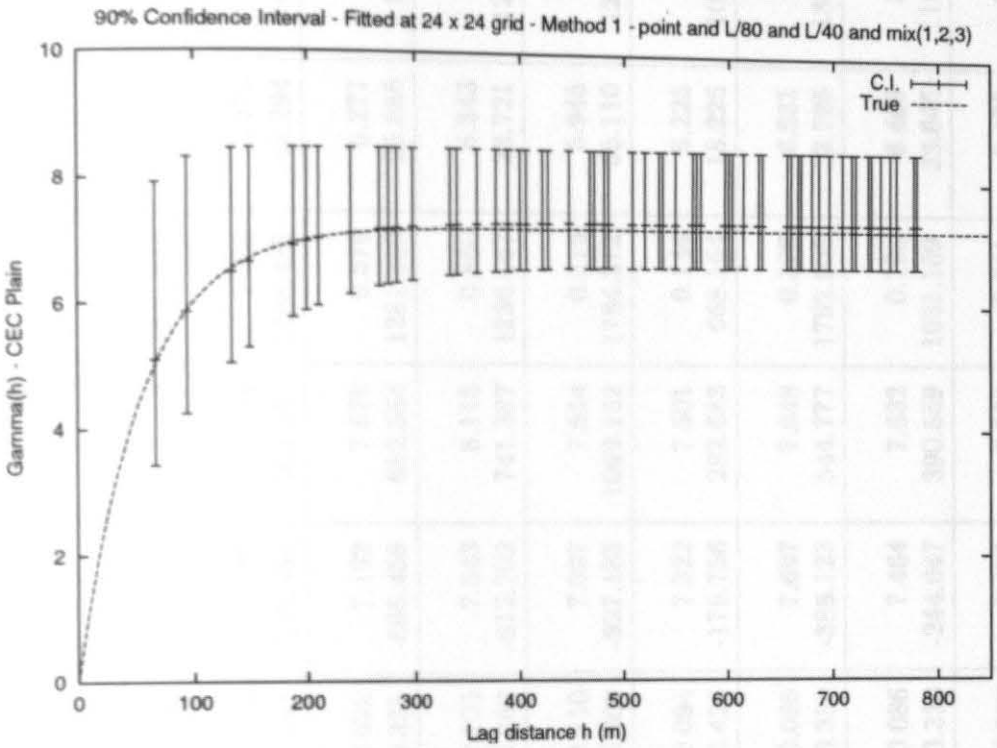


Figure D.9: Numerical Experiment 2 Plain, (192 points) - point and L/80 and L/40



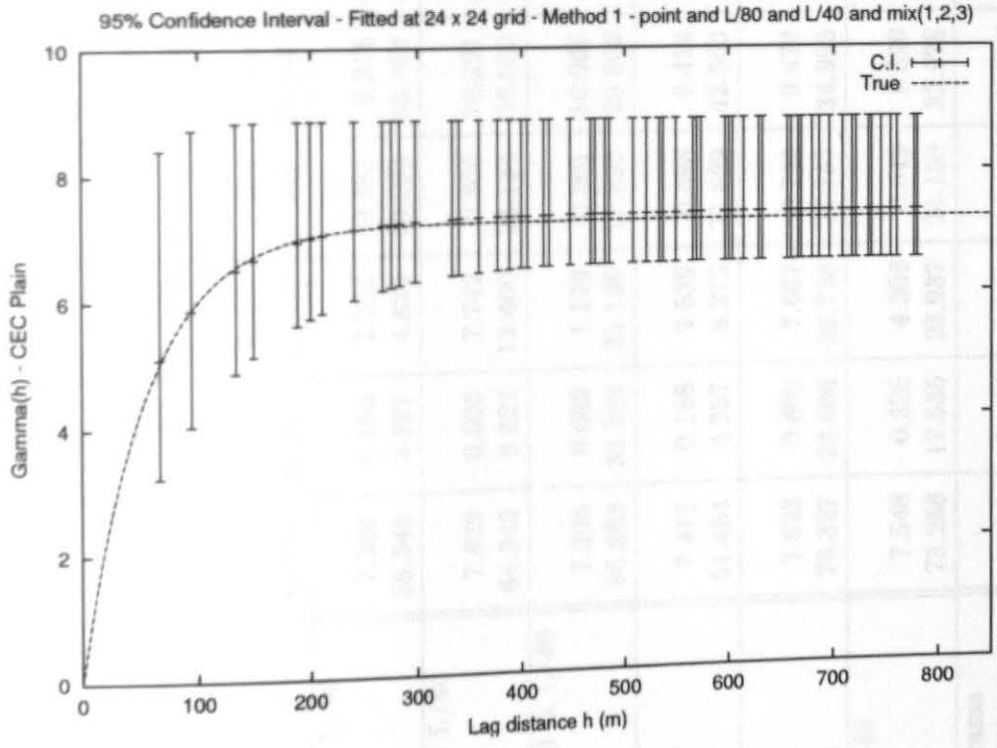


Figure D.10: Numerical Experiment 2 Plain, (192 points) - all variograms at once

Parameter	Mean	Bias	%Bias	S.D.	%S.D.	C.V.	Low C.I.	High C.I.	S.D. C.I.	Low C.I.	High C.I.
Point:											
Sill	7.459	0.236	3.165	0.838	11.234	0.112	7.268	7.650	0.382	6.004	9.451
Range	59.917	4.196	7.002	14.900	24.868	0.249	-811.008	930.841	1328.040	24.267	89.806
L/80:											
Sill	6.984	0.239	3.420	0.855	12.237	0.122	6.622	7.346	1.365	5.595	8.823
Range	67.276	11.555	17.176	17.462	25.956	0.260	-873.697	1008.250	1741.792	36.622	106.125
L/40:											
Sill	7.890	0.667	8.457	1.294	16.404	0.164	7.395	8.385	1.798	5.470	10.303
Range	86.535	30.814	35.608	23.278	26.900	0.269	-730.488	903.557	1388.487	43.294	141.678
Mixed: pt & L/80											
Sill	7.386	0.163	2.202	0.681	9.216	0.092	7.192	7.579	0.579	6.277	8.690
Range	58.548	2.827	4.829	19.593	33.464	0.335	-565.458	682.554	1281.859	25.855	96.073
Mixed: pt & L/40											
Sill	7.829	0.606	7.742	0.806	10.291	0.103	7.543	8.115	0.860	6.343	9.428
Range	64.343	8.622	13.400	23.184	36.032	0.360	-612.702	741.387	1236.567	23.721	122.721
Mixed: L/80 & L/40											
Sill	7.305	0.082	1.129	0.801	10.966	0.110	7.097	7.514	0.515	5.945	8.830
Range	85.983	30.262	35.196	17.893	20.809	0.208	-927.185	1099.152	1784.312	55.110	123.844
pt + L/80											
Sill	7.411	0.188	2.539	0.698	9.414	0.094	7.322	7.501	0.144	6.225	8.824
Range	51.464	4.257	8.272	21.899	42.553	0.426	-179.756	282.683	668.164	18.225	100.927
pt + L/40											
Sill	7.823	0.600	7.667	0.741	9.474	0.095	7.697	7.948	0.376	6.581	9.337
Range	79.327	23.606	29.758	27.761	34.995	0.350	-386.123	544.777	1792.605	32.795	135.757
L/80 + L/40											
Sill	7.548	0.325	4.309	0.649	8.599	0.086	7.464	7.632	0.193	6.409	8.898
Range	73.256	17.535	23.937	24.159	32.978	0.330	-244.047	390.559	1031.109	33.646	116.459
All variograms											
Sill	7.530	0.307	4.079	0.622	8.257	0.083	7.499	7.562	0.041	6.466	8.736
Range	69.887	14.166	20.270	19.433	27.806	0.278	-45.603	185.377	221.400	34.720	102.372

Table D.2: Experiment 3 Plain, 12x12, Sill = 7.223, Range = 3*55.721. Confidence Intervals (C.I.) are at the 95% level

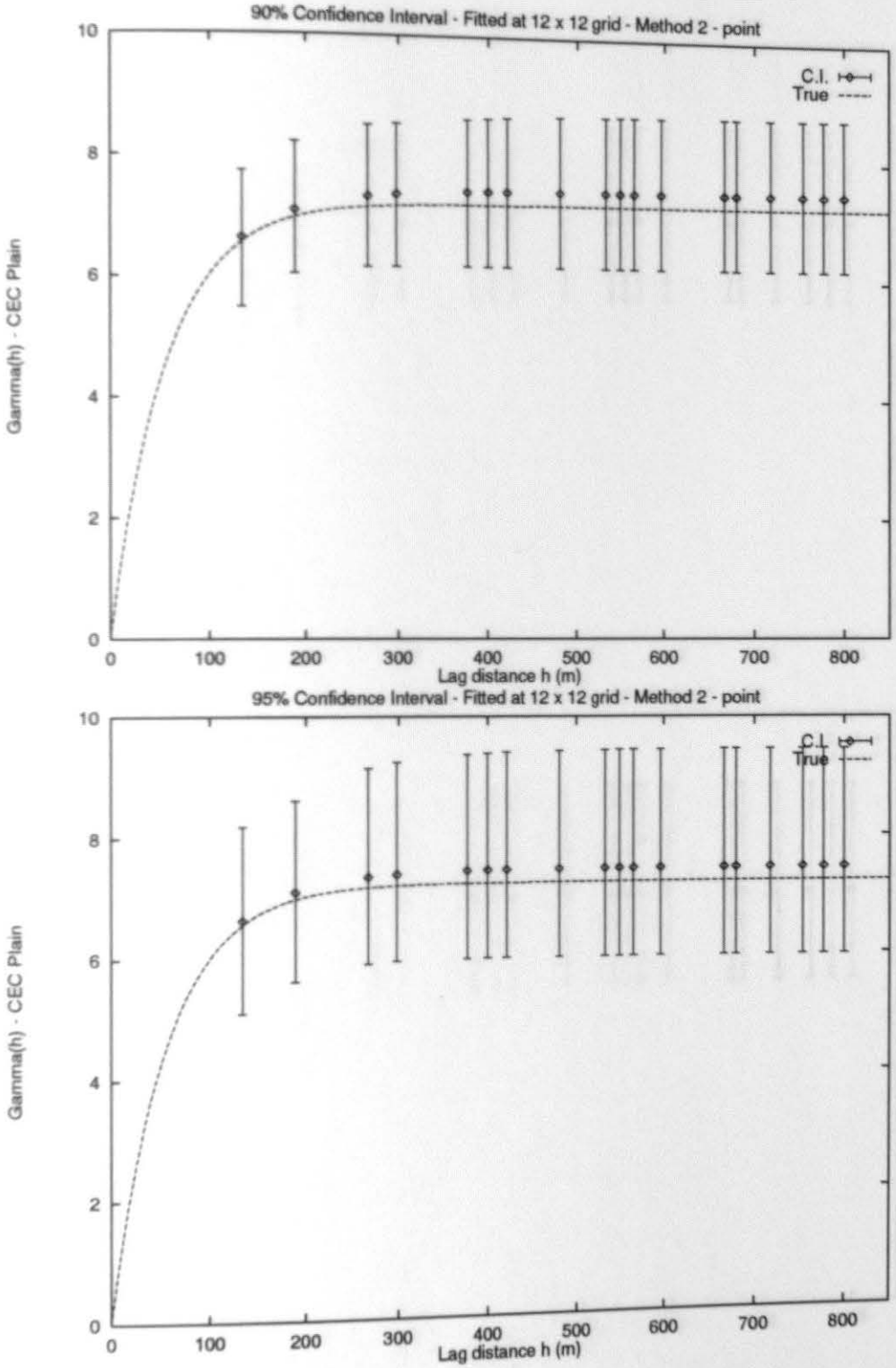


Figure D.11: Numerical Experiment 3 Plain, (192 points) - point

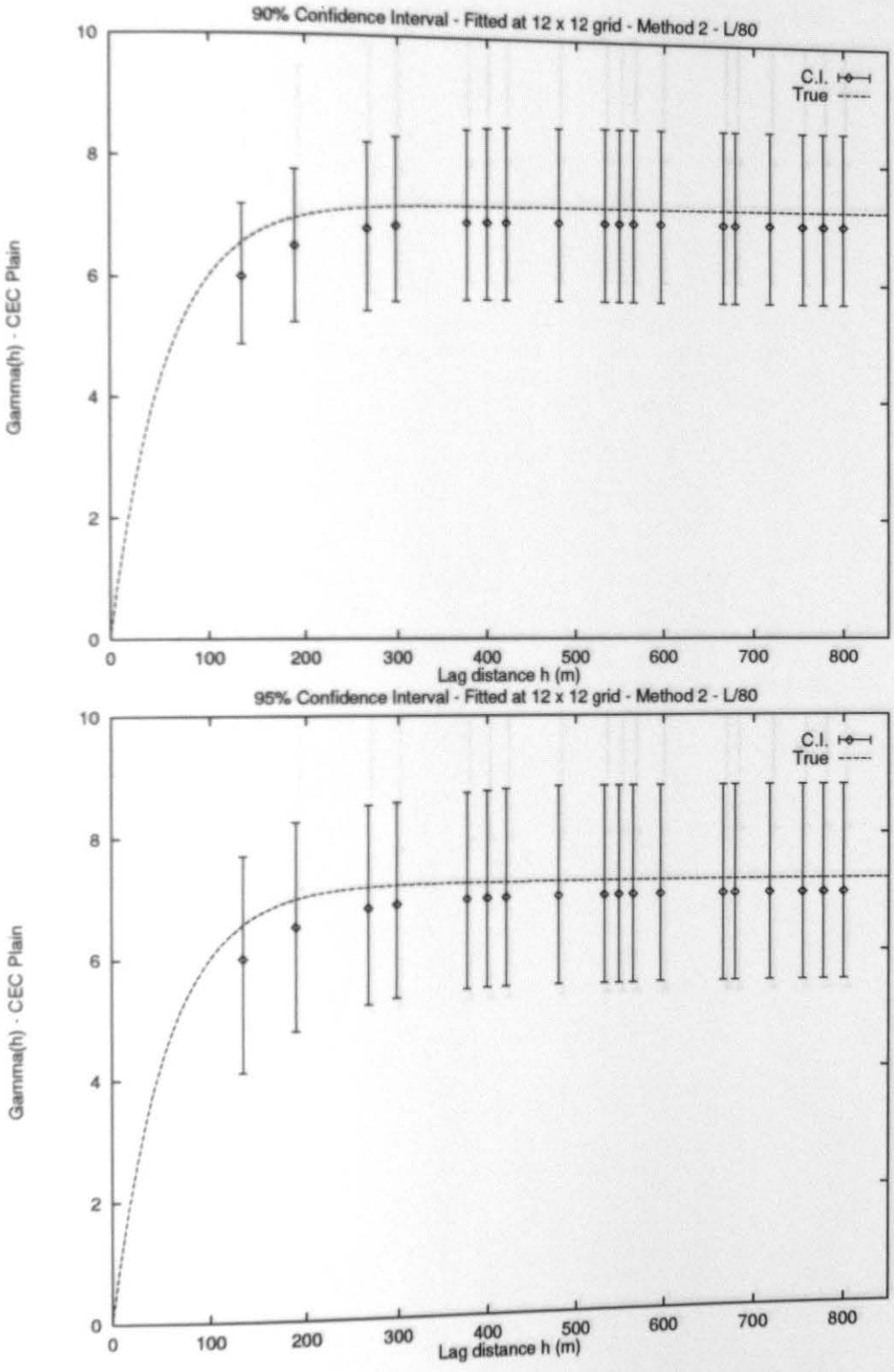


Figure D.12: Numerical Experiment 3 Plain, (192 points) - L/80

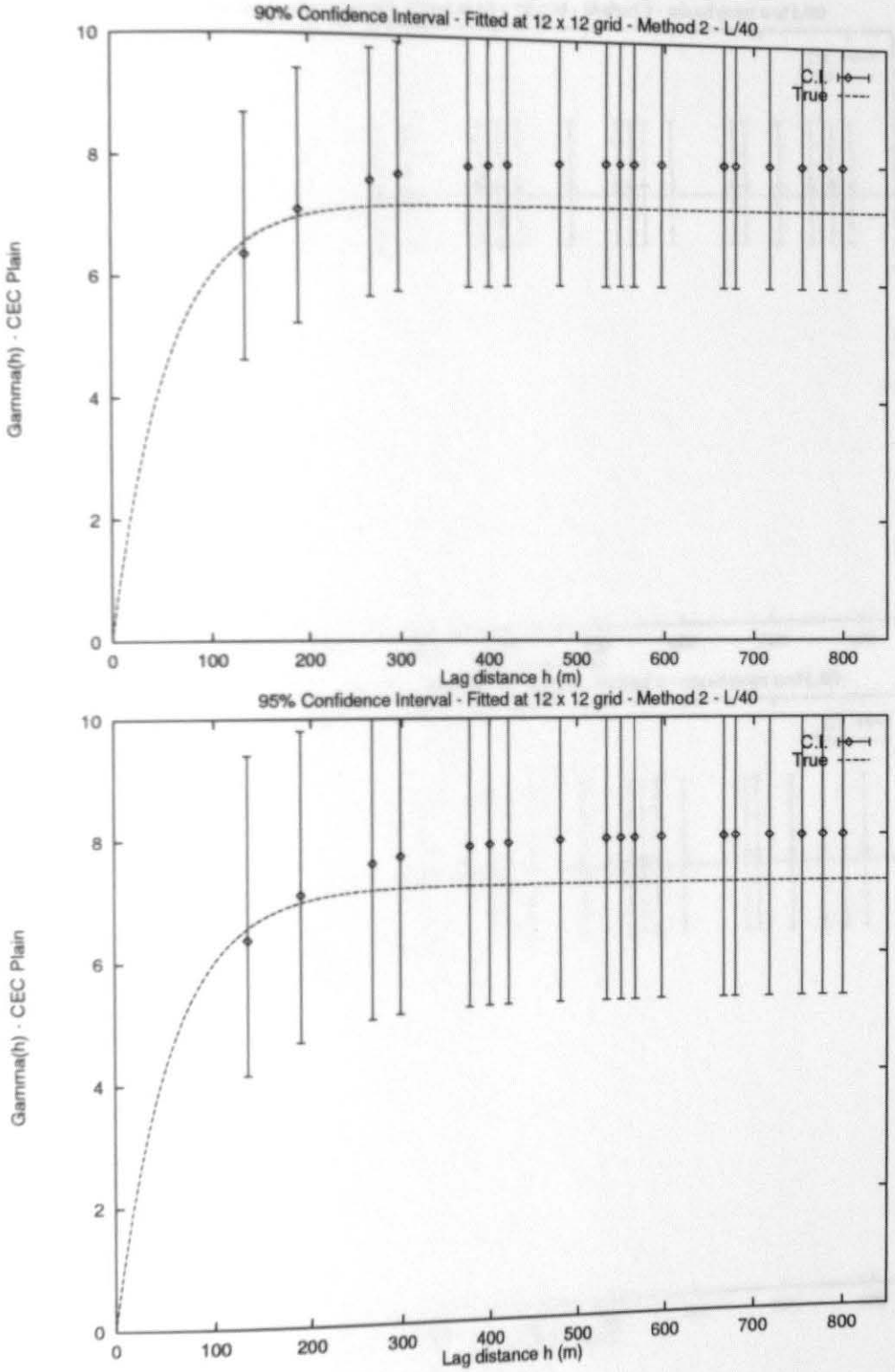


Figure D.13: Numerical Experiment 3 Plain, (192 points) - L/40

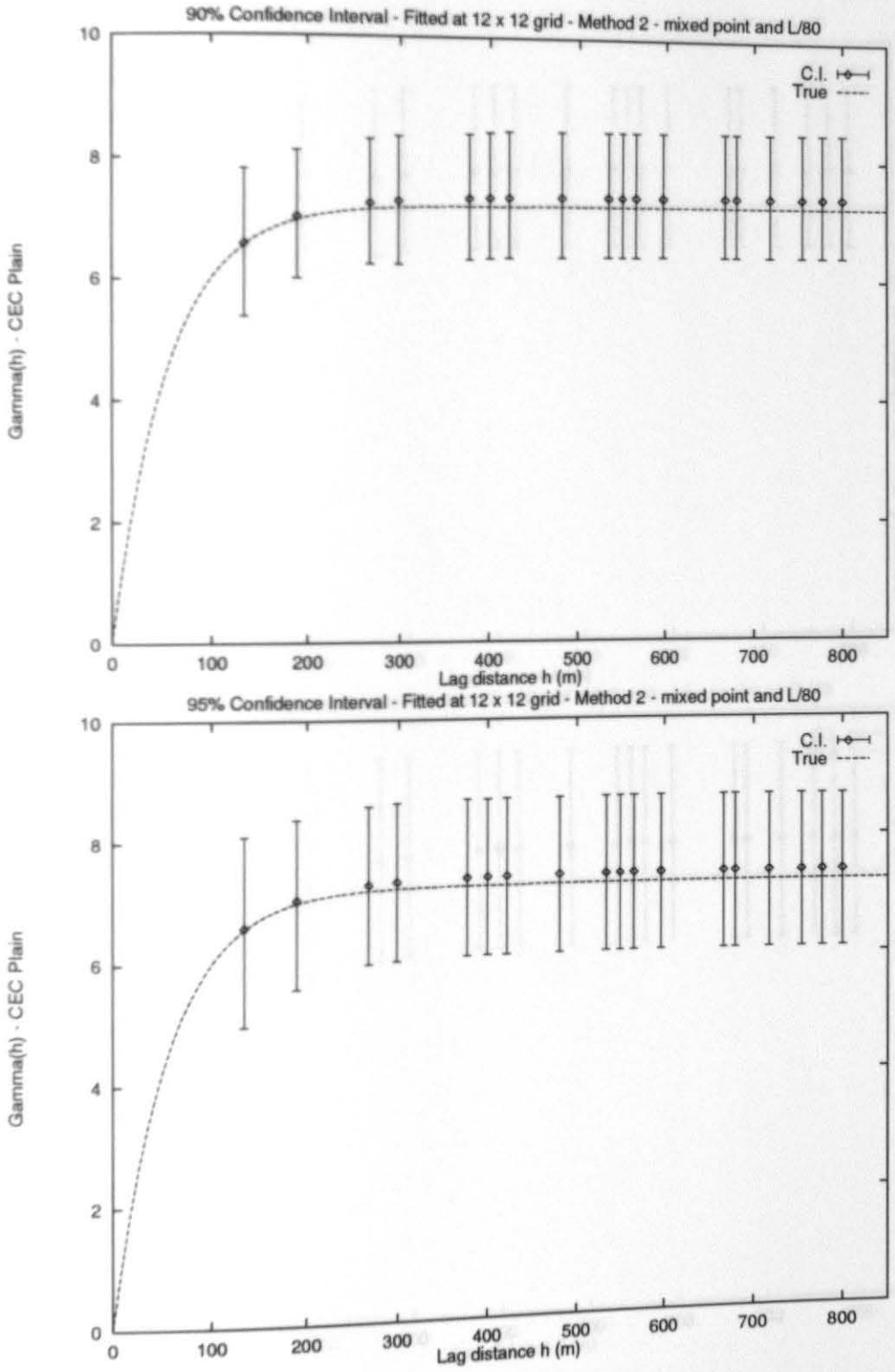


Figure D.14: Numerical Experiment 3 Plain, (192 points) - mixed point and L/80

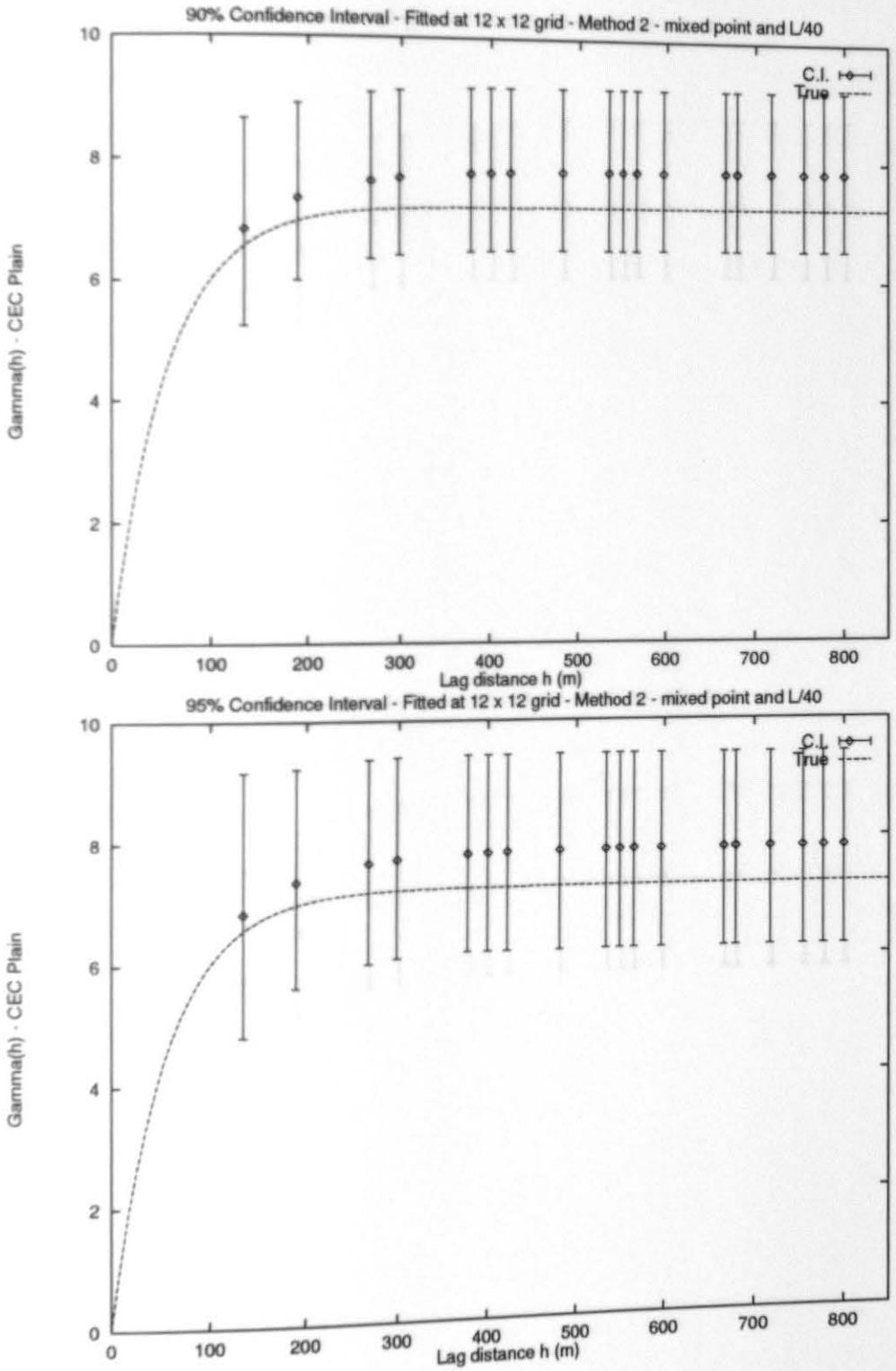


Figure D.15: Numerical Experiment 3 Plain, (192 points) - mixed point and L/40

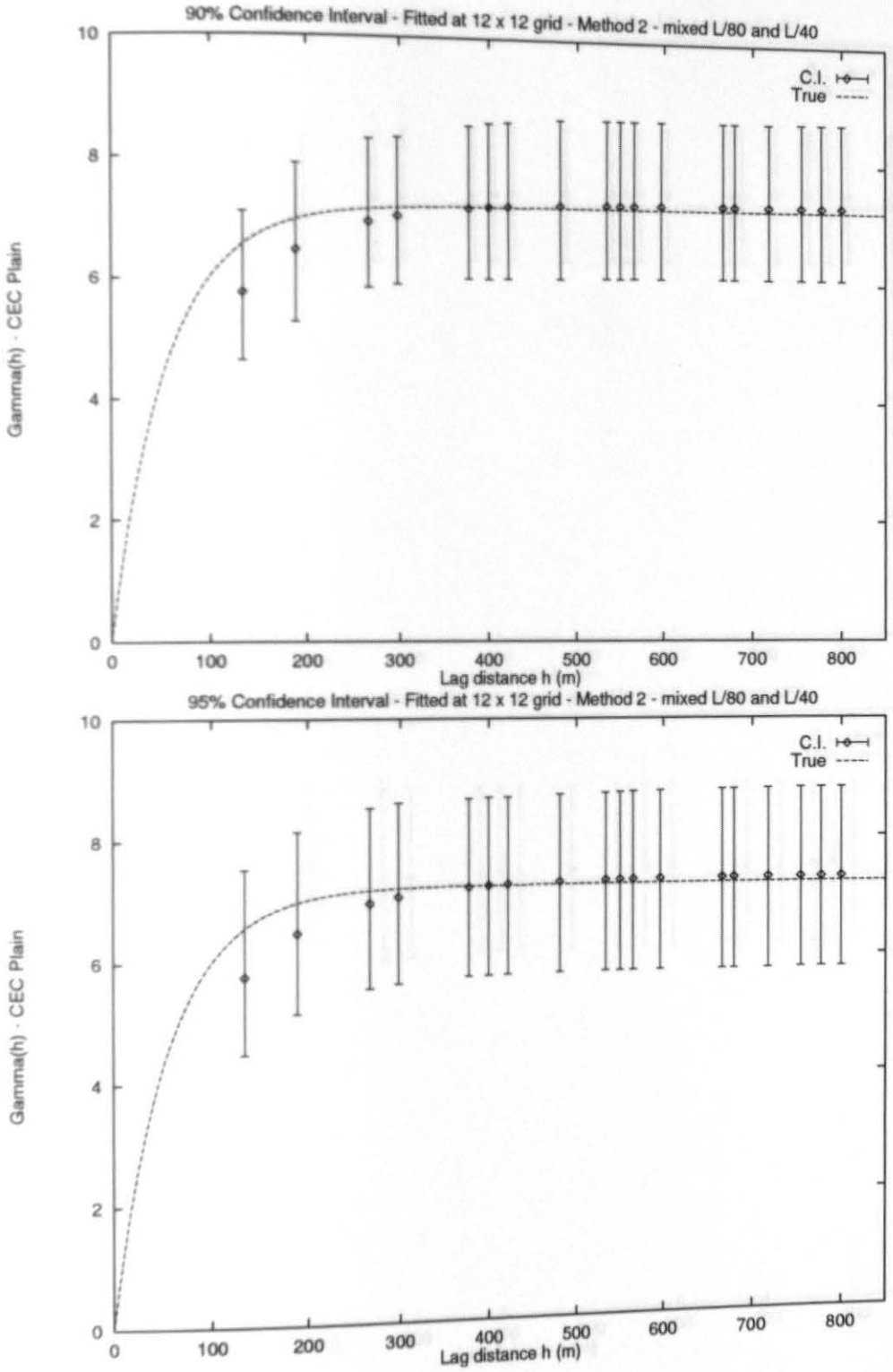


Figure D.16: Numerical Experiment 3 Plain, (192 points) - mixed L/80 and L/40

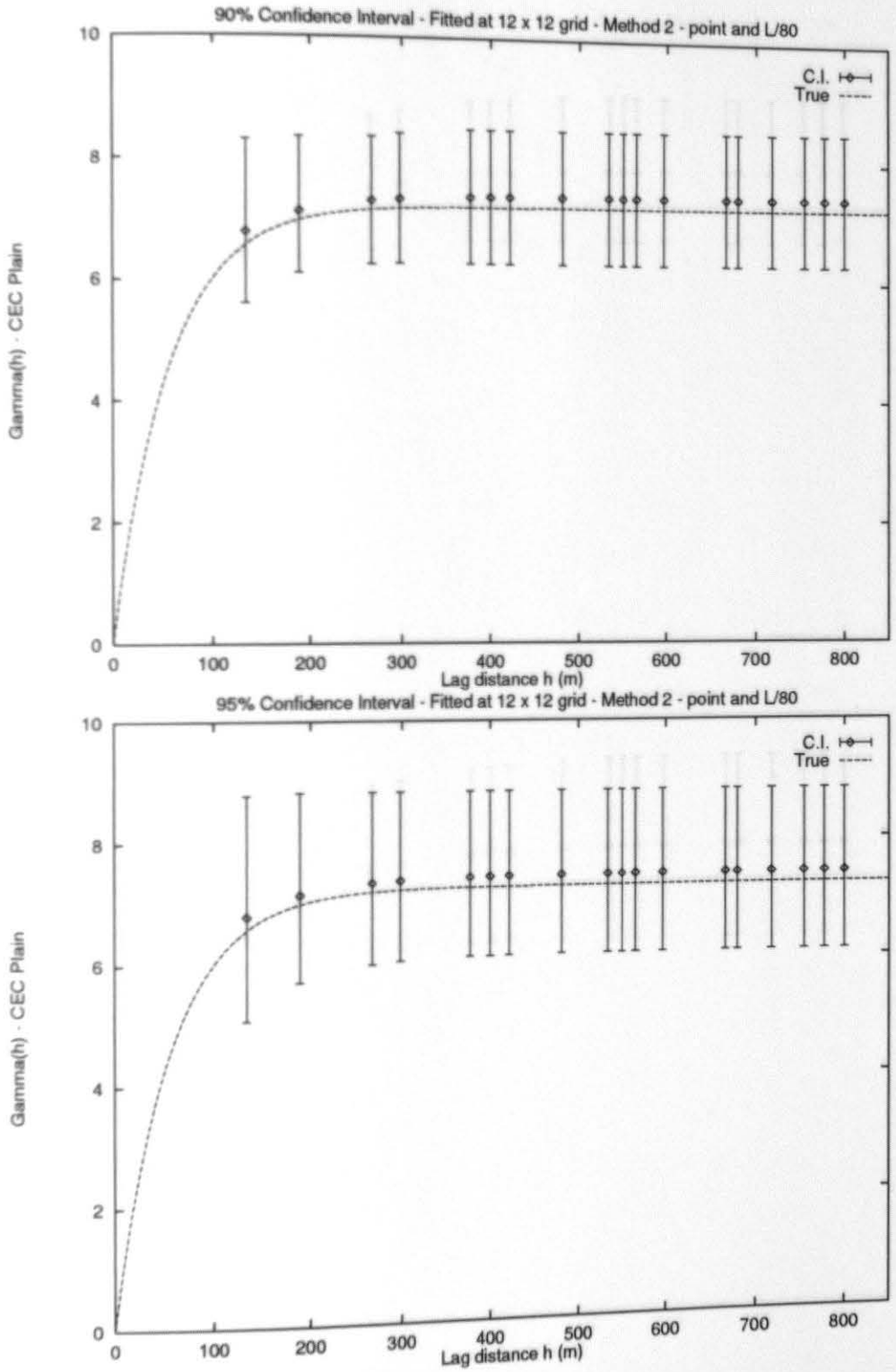


Figure D.17: Numerical Experiment 3 Plain, (192 points) - point and L/80

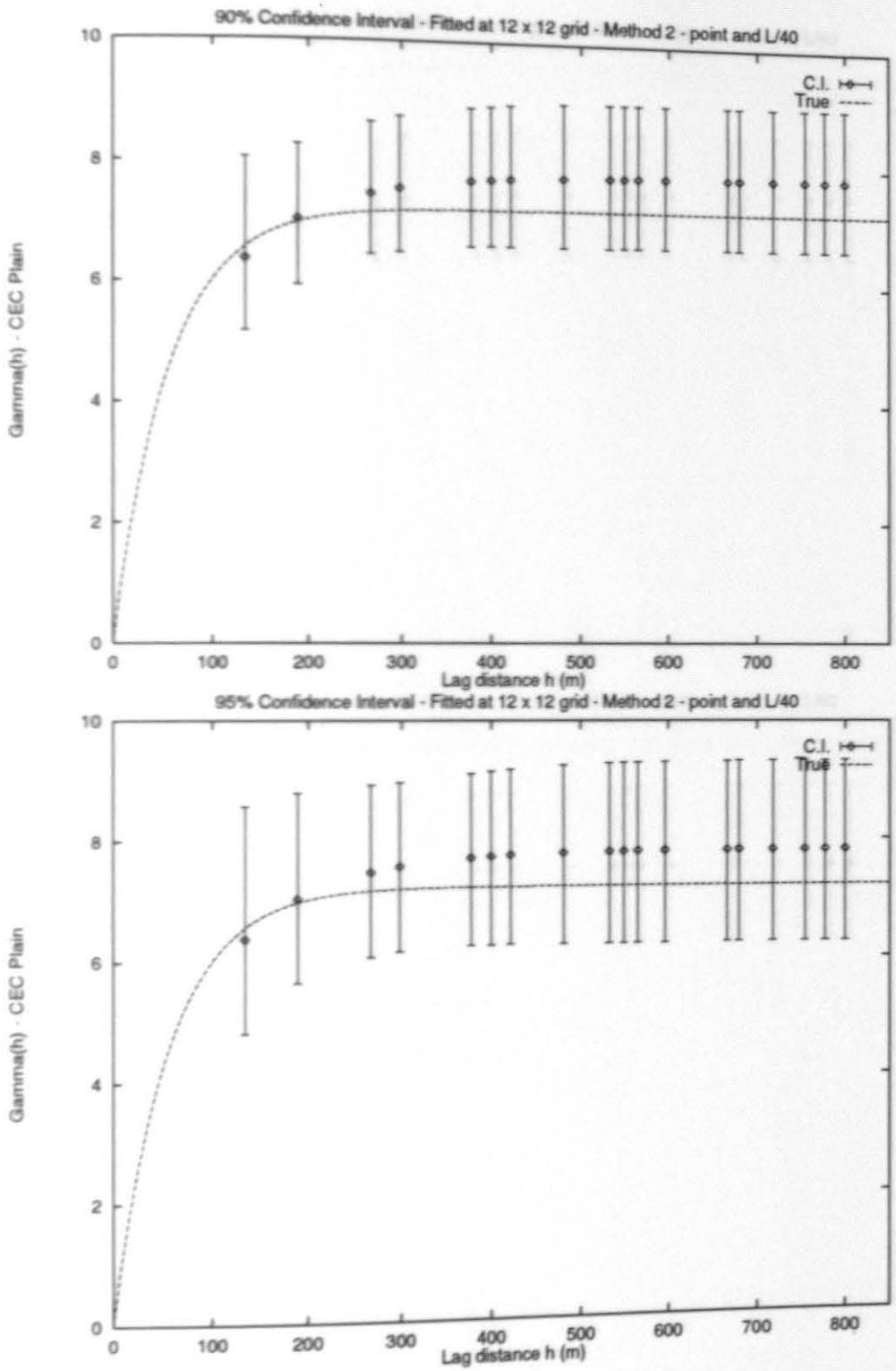


Figure D.18: Numerical Experiment 3 Plain, (192 points) - point and L/40

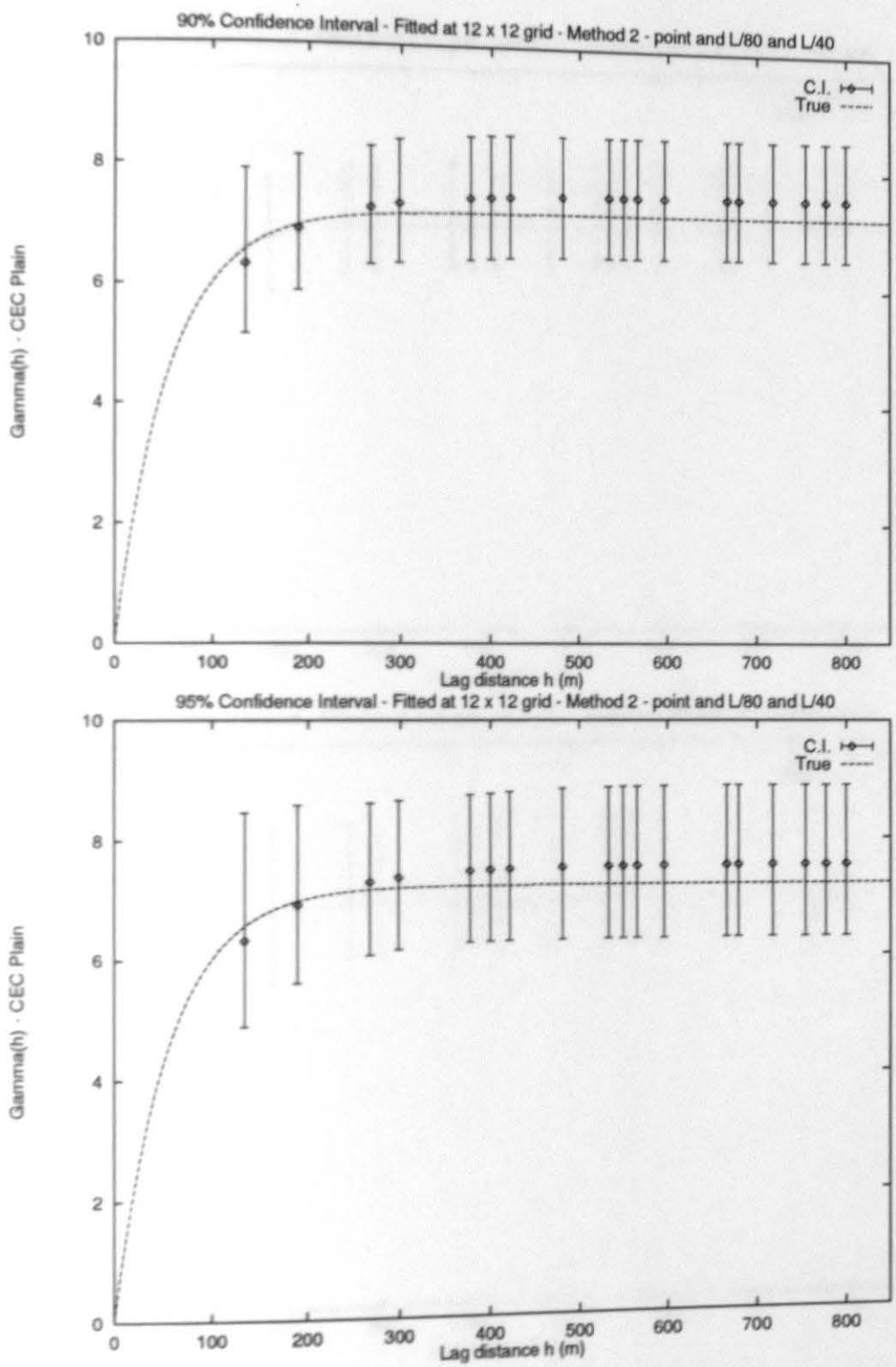


Figure D.19: Numerical Experiment 3 Plain, (192 points) - point and L/80 and L/40

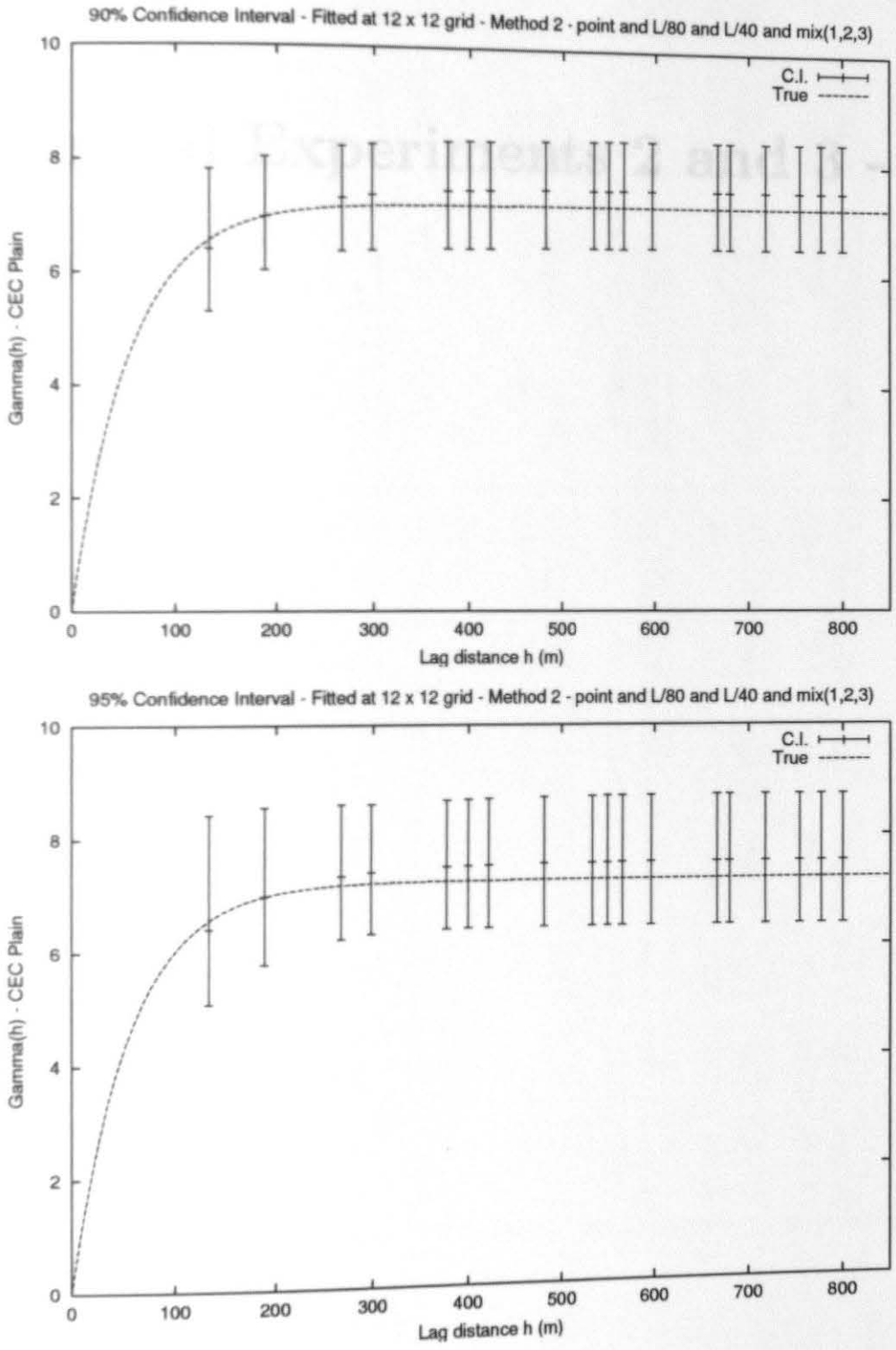


Figure D.20: Numerical Experiment 3 Plain, (192 points) - all variograms at once

Numerical Experiments 2 and 3 - Nugget

Parameter	Mean	Bias	%Bias	S.D.	%S.D.	C.V.	Low C.I.	High C.I.	S.D. C.I.	Low C.I.	High C.I.
Point:											
Sill	5.600	0.089	1.596	1.136	20.285	0.203	5.476	5.723	0.299	3.460	8.032
Range	63.042	0.858	1.361	21.258	33.721	0.337	-694.303	820.388	1345.028	27.269	121.674
L/80:											
Sill	5.641	0.048	0.842	0.952	16.869	0.169	5.096	6.187	9.612	4.218	8.341
Range	67.619	5.435	8.037	18.467	27.311	0.273	-281.972	417.209	942.934	43.055	121.220
L/40:											
Sill	5.595	0.094	1.674	0.884	15.790	0.158	5.428	5.762	0.763	4.271	7.724
Range	66.375	4.191	6.314	17.079	25.731	0.257	-178.892	311.641	720.376	38.814	114.475
Mixed: pt & L/80											
Sill	5.584	0.105	1.883	0.625	11.197	0.112	5.492	5.676	0.352	4.540	6.962
Range	65.985	3.801	5.760	21.789	33.021	0.330	-467.283	599.253	844.569	32.423	127.037
Mixed: pt & L/40											
Sill	5.541	0.148	2.680	0.690	12.448	0.124	5.450	5.631	0.248	4.279	7.069
Range	66.698	4.514	6.768	19.485	29.214	0.292	-425.070	558.466	783.191	35.384	110.887
Mixed: L/80 & L/40											
Sill	5.558	0.131	2.350	0.508	9.138	0.091	5.452	5.664	0.382	4.775	6.669
Range	66.482	4.298	6.465	17.399	26.171	0.262	-233.037	366.001	677.326	43.457	122.608
pt + L/80											
Sill	5.813	0.124	2.128	0.778	13.379	0.134	5.733	5.893	0.257	4.569	7.739
Range	60.527	1.657	2.738	31.550	52.125	0.521	-411.371	532.425	2022.688	14.313	119.535
pt + L/40											
Sill	5.741	0.052	0.901	0.841	14.643	0.146	5.691	5.790	0.125	4.283	7.741
Range	65.322	3.138	4.804	28.486	43.608	0.436	-205.601	336.245	1443.540	22.206	133.861
L/80 + L/40											
Sill	5.746	0.057	0.991	0.648	11.273	0.113	5.699	5.793	0.094	4.911	7.647
Range	64.556	2.372	3.674	26.291	40.727	0.407	-208.231	337.342	1134.503	21.681	120.023
All variograms											
Sill	5.689	0.000	0.004	0.607	10.669	0.107	5.670	5.708	0.032	4.885	7.609
Range	62.379	0.195	0.313	21.884	35.083	0.351	-28.046	152.804	255.041	21.976	106.883

Table E.1: Experiment 2 Nugget, 24x24, Sill = 5.689, Range = 3*62.184. Confidence Intervals (C.I.) are at the 95% level

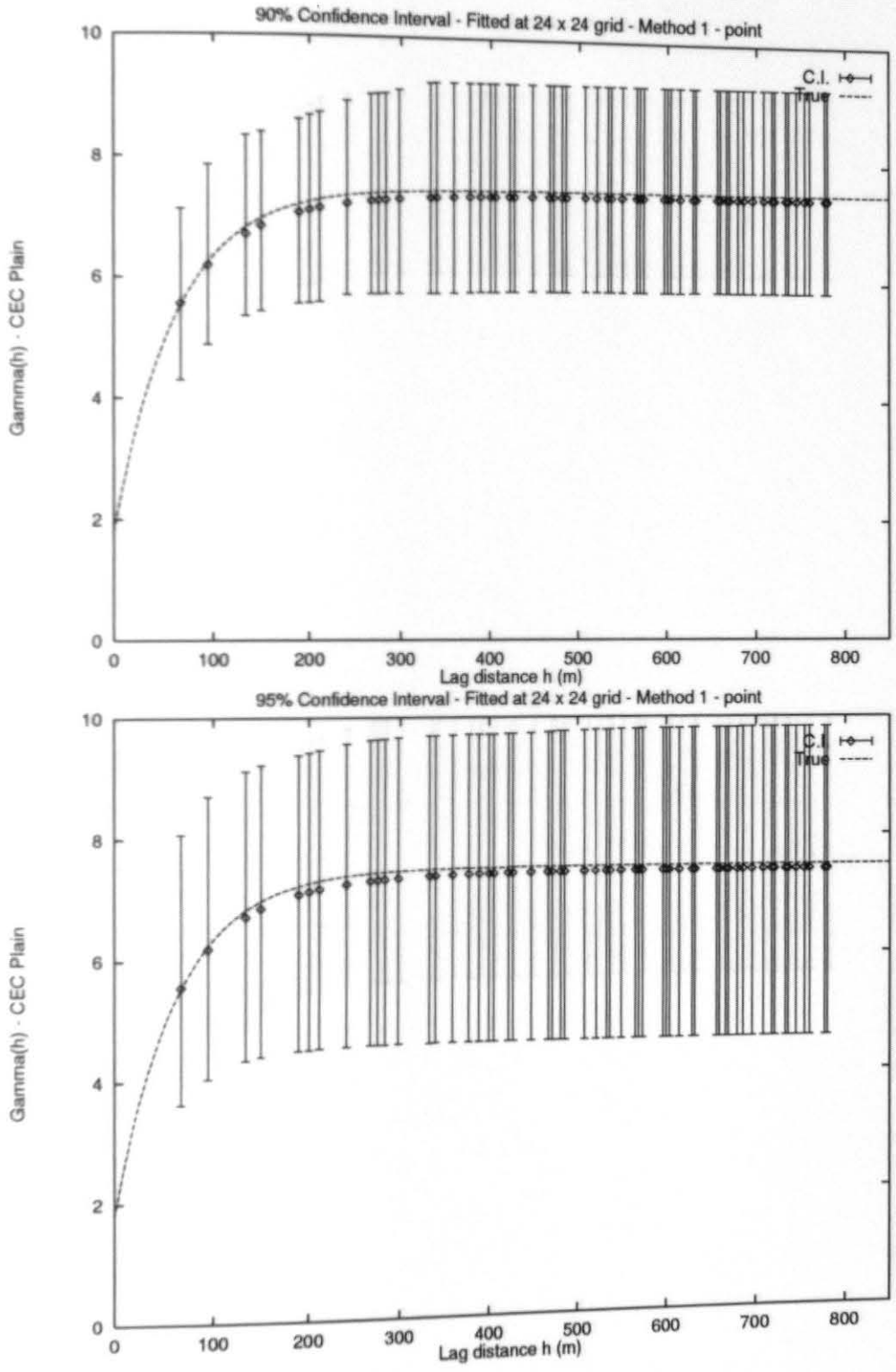


Figure E.1: Numerical Experiment 2 Nugget, (192 points) - Point

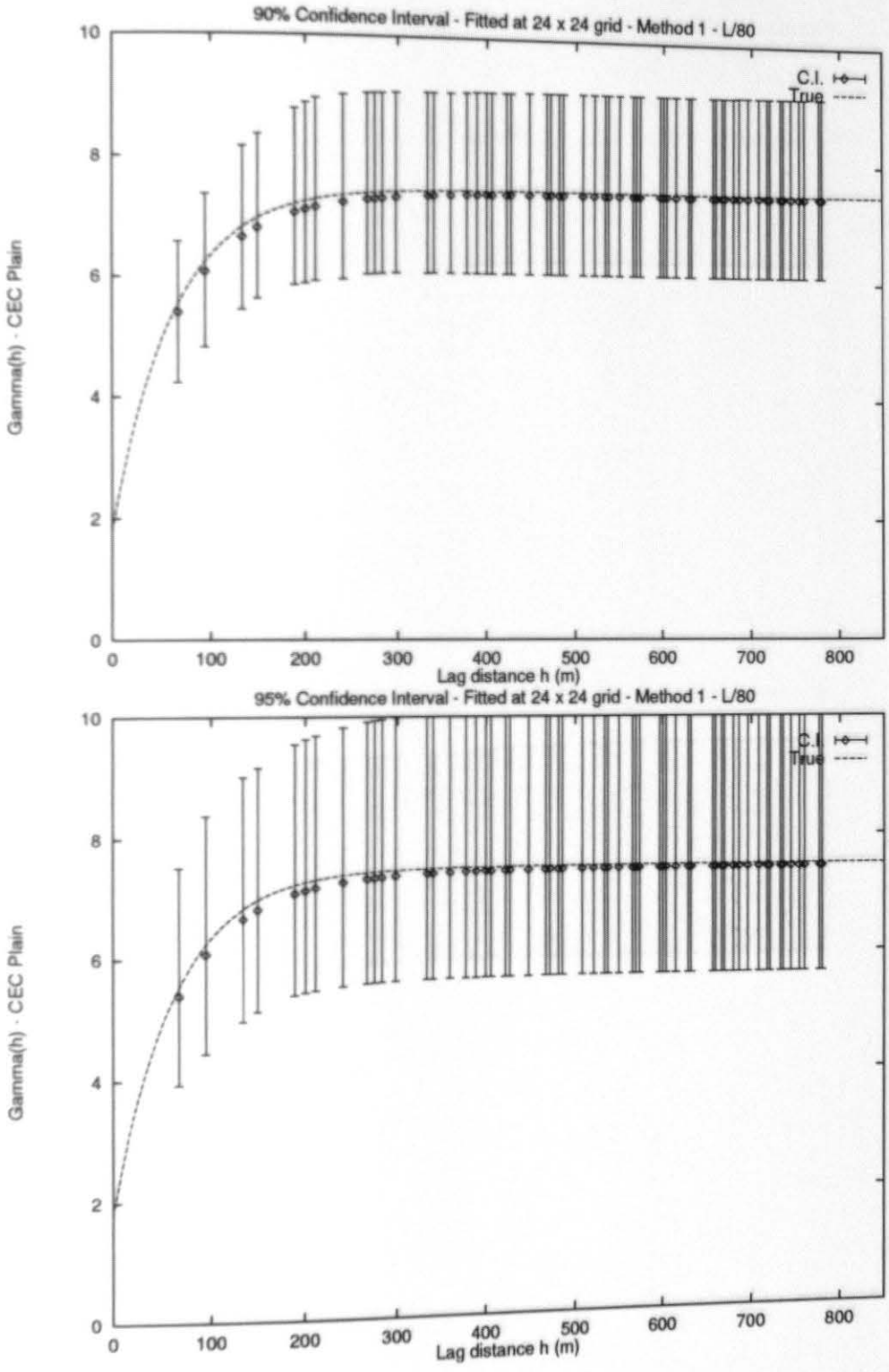


Figure E.2: Numerical Experiment 2 Nugget, (192 points) - L/80

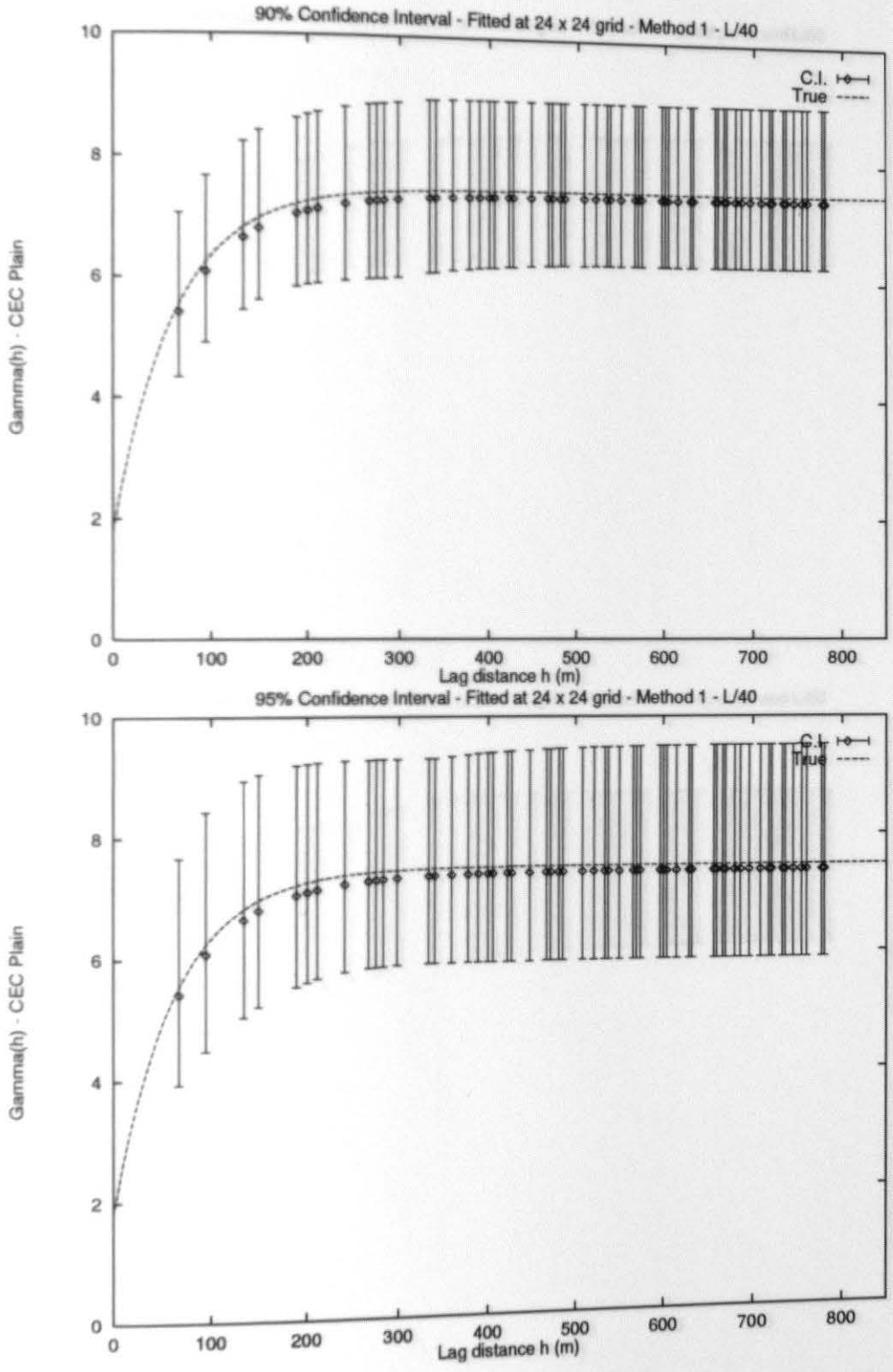


Figure E.3: Numerical Experiment 2 Nugget, (192 points) - L/40

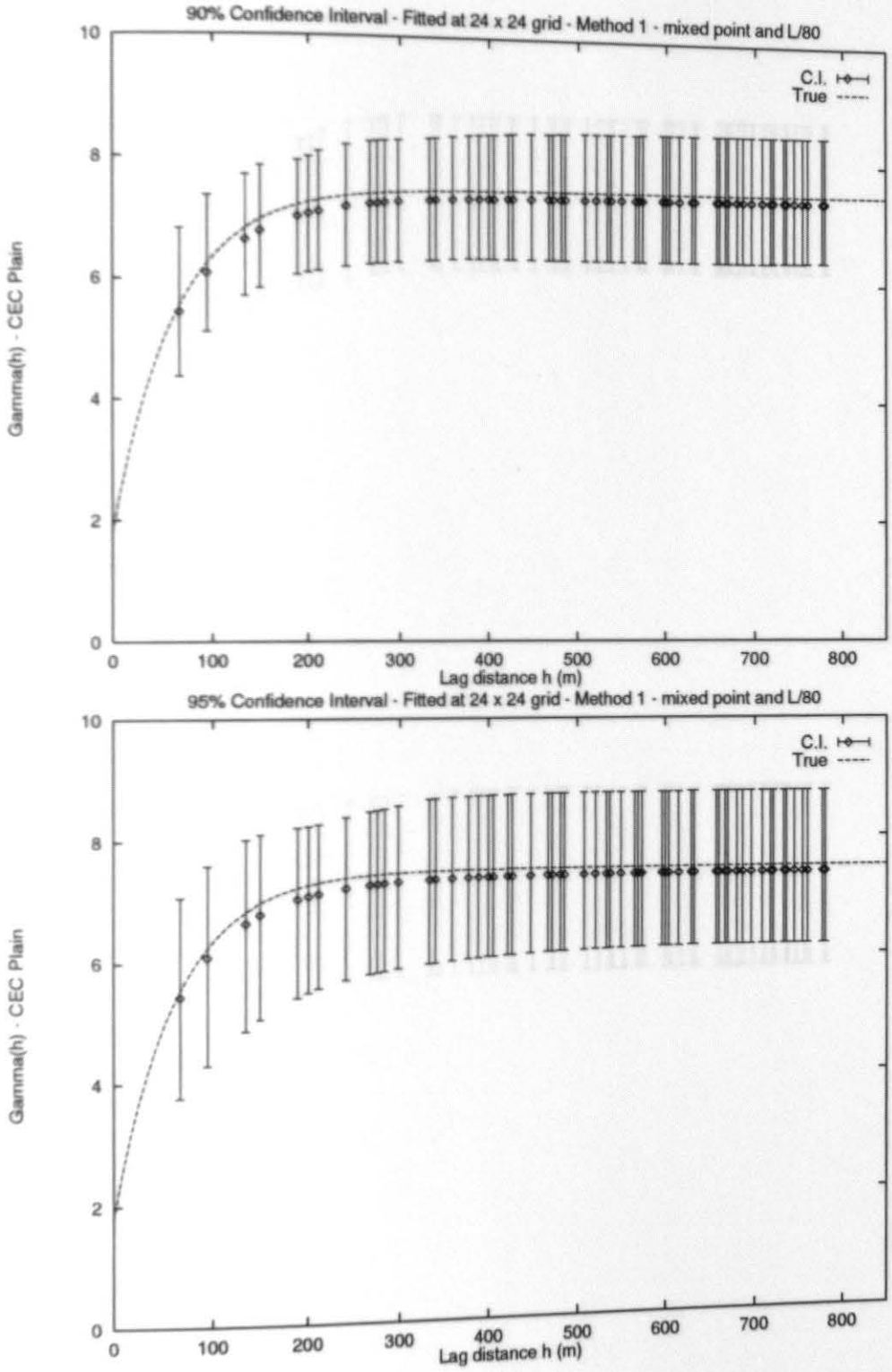


Figure E.4: Numerical Experiment 2 Nugget, (192 points) - mixed point and L/80

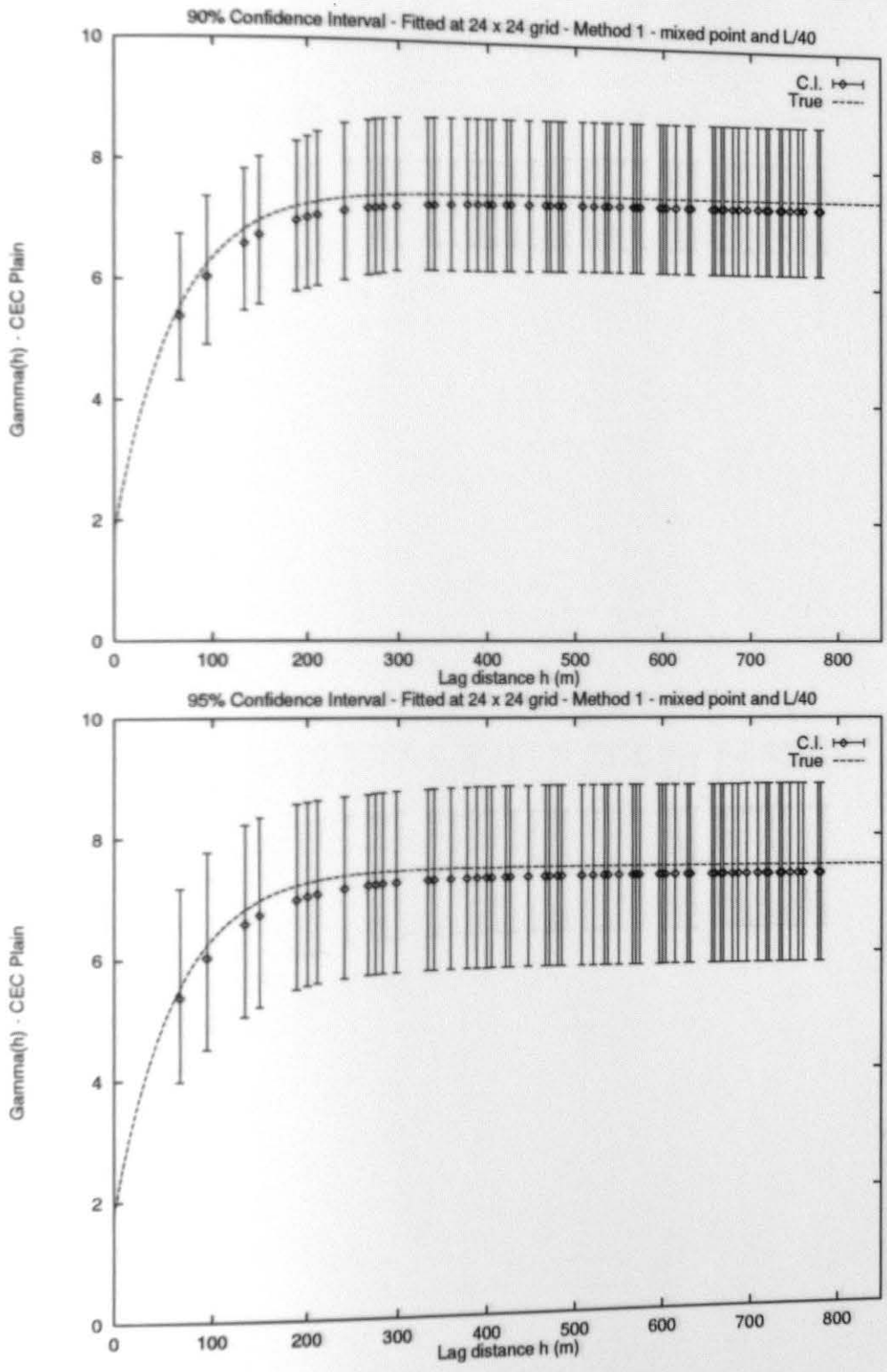


Figure E.5: Numerical Experiment 2 Nugget, (192 points) - mixed point and L/40

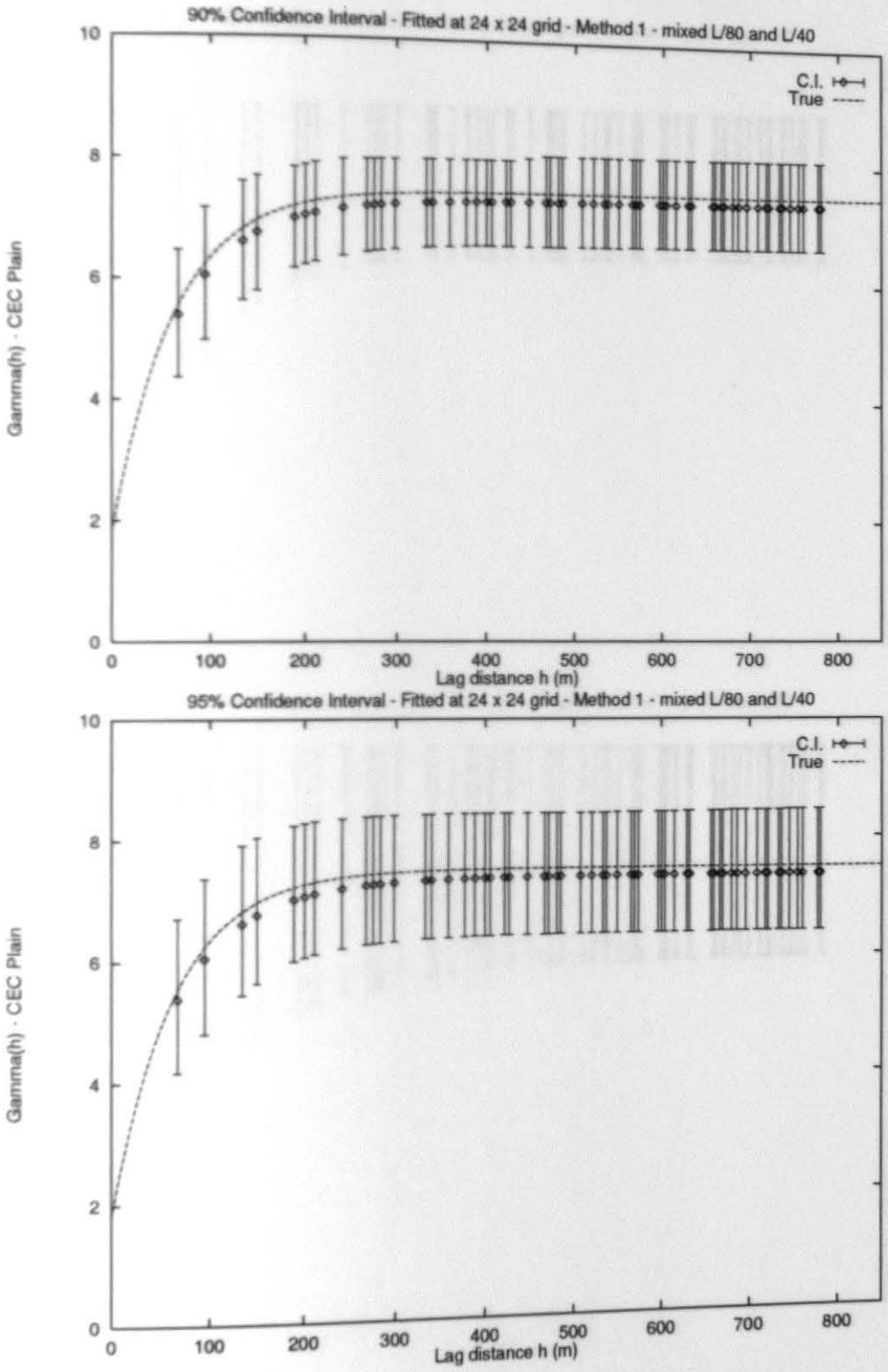


Figure E.6: Numerical Experiment 2 Nugget, (192 points) - mixed L/80 and L/40

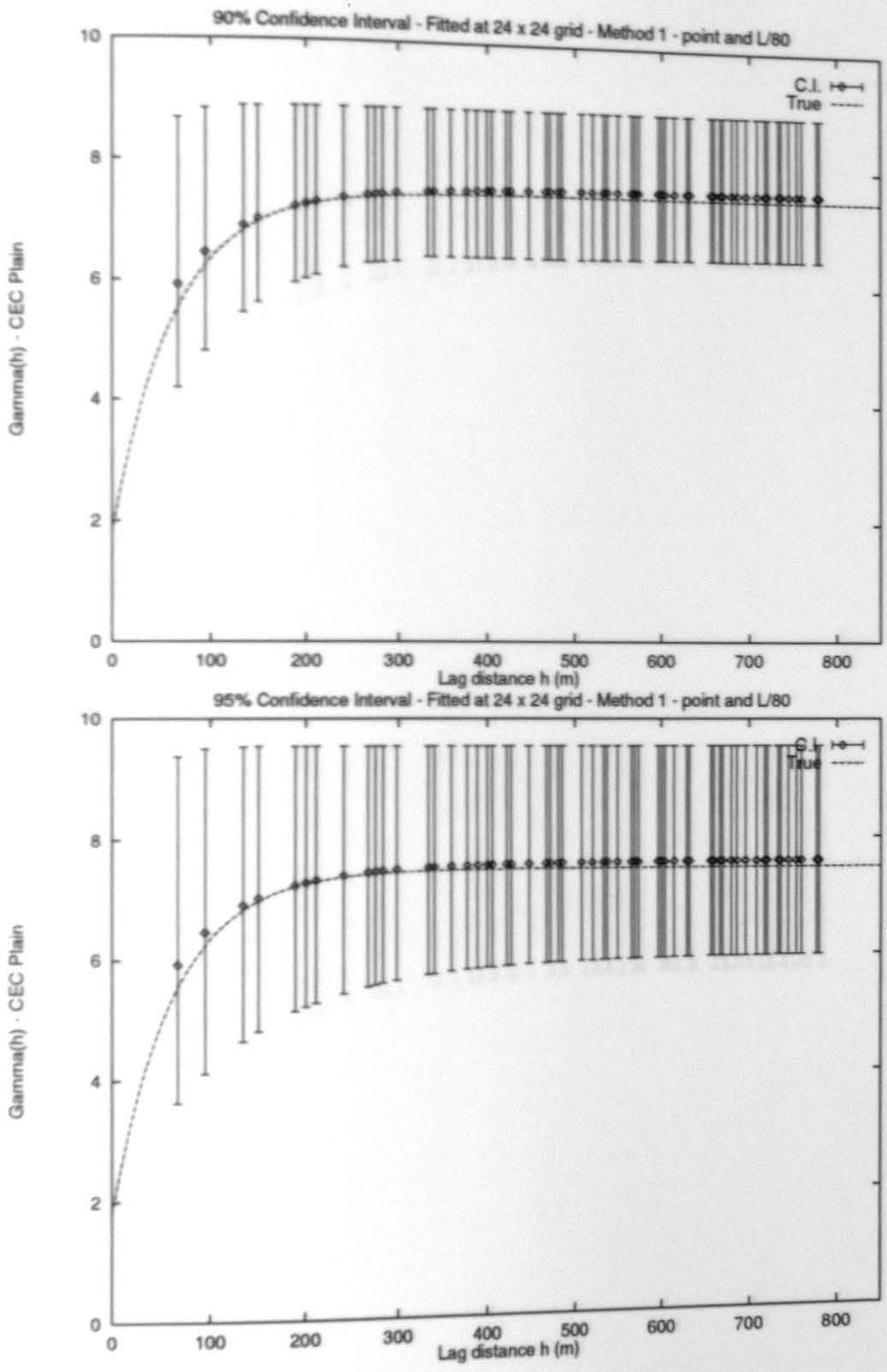


Figure E.7: Numerical Experiment 2 Nugget, (192 points) - point and L/80

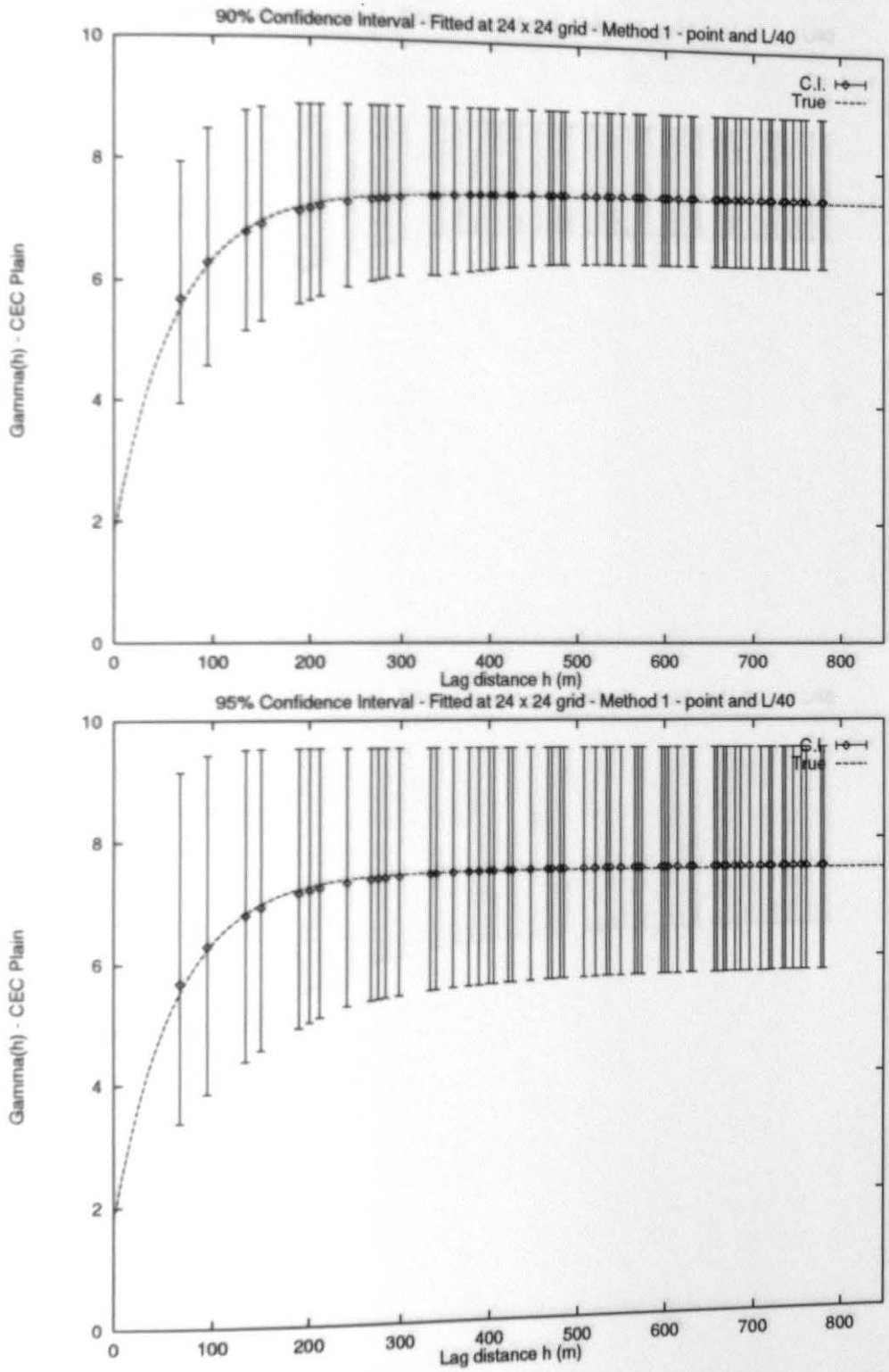


Figure E.8: Numerical Experiment 2 Nugget, (192 points) - point and L/40

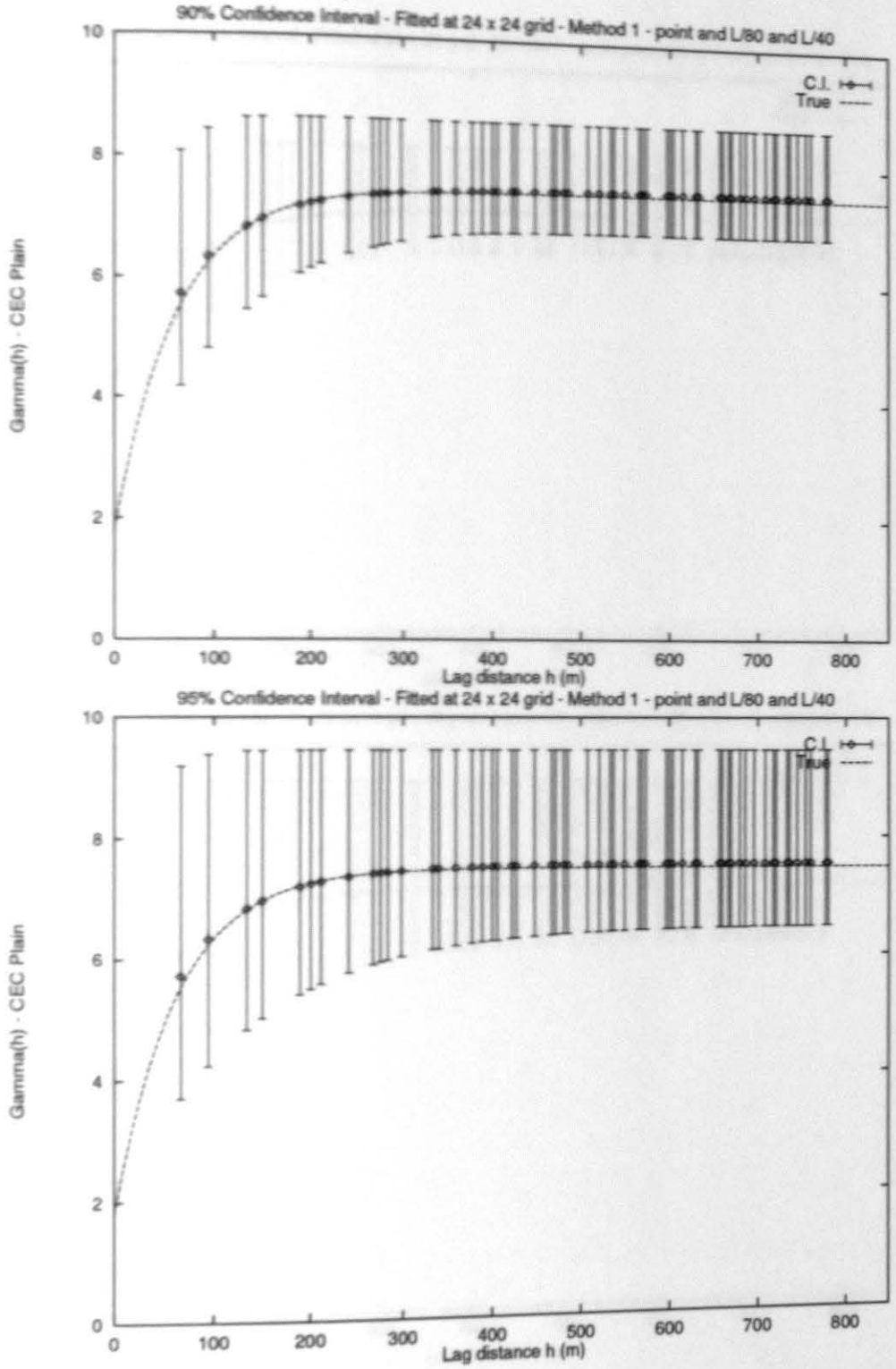


Figure E.9: Numerical Experiment 2 Nugget, (192 points) - point and L/80 and L/40

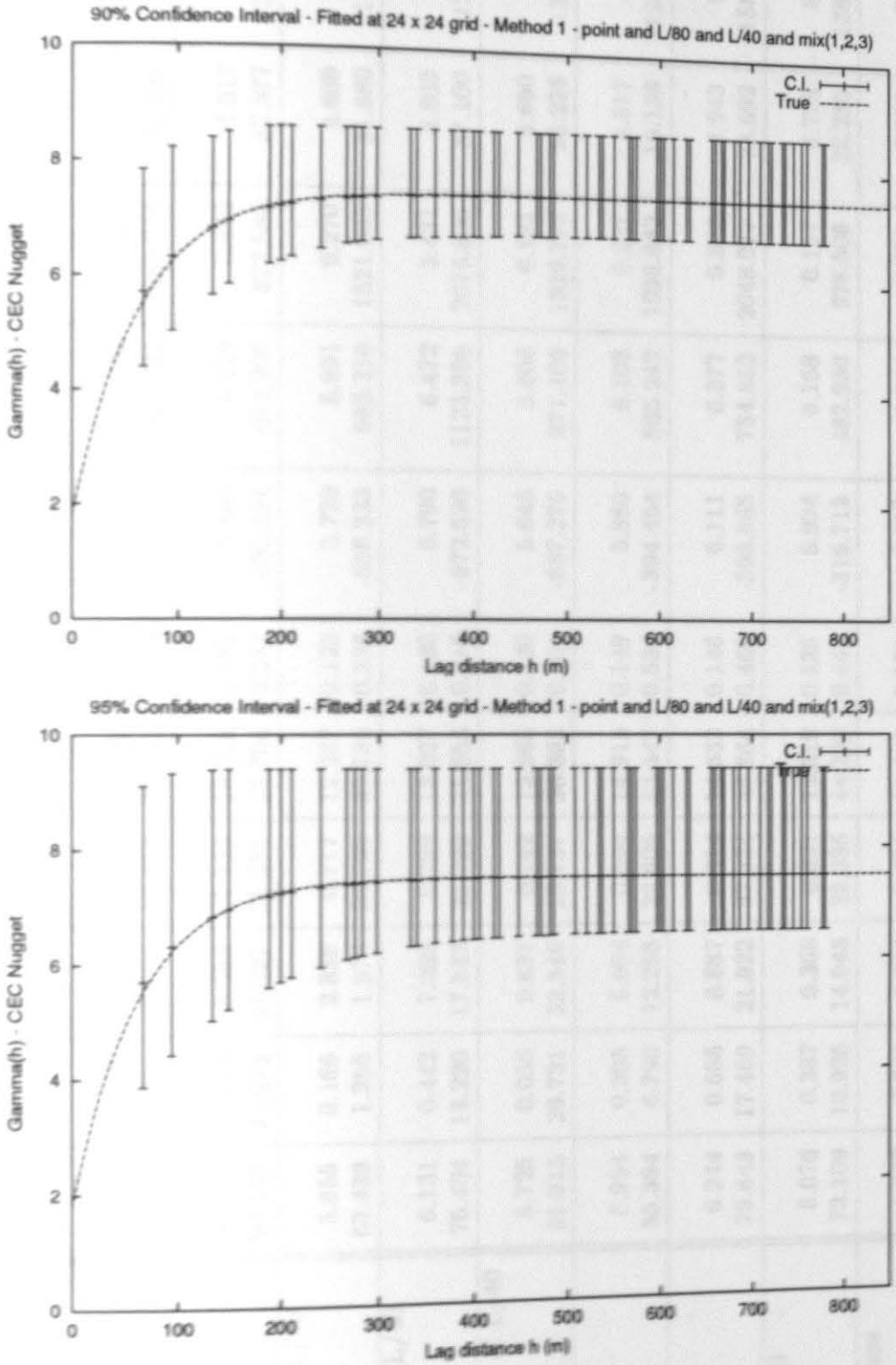


Figure E.10: Numerical Experiment 2 Nugget, (192 points) - all variograms at once

Parameter	Mean	Bias	%Bias	S.D.	%S.D.	C.V.	Low C.I.	High C.I.	S.D. C.I.	Low C.I.	High C.I.
Point:											
Sill	6.008	0.319	5.313	1.084	18.047	0.180	5.703	6.313	0.480	3.884	8.529
Range	69.711	7.527	10.797	26.925	38.624	0.386	-1930.855	2070.276	2311.211	25.992	131.314
L/80:											
Sill	5.511	0.178	3.233	0.795	14.426	0.144	5.292	5.730	1.253	4.271	7.029
Range	73.252	11.068	15.110	17.488	23.873	0.239	-781.226	927.731	1883.746	35.585	108.834
L/40:											
Sill	6.007	0.318	5.293	1.125	18.723	0.187	5.885	6.129	0.470	4.517	9.292
Range	90.157	27.973	31.027	21.375	23.708	0.237	-432.591	612.906	822.594	57.577	132.239
Mixed: pt & L/80											
Sill	5.855	0.166	2.838	0.717	12.247	0.122	5.720	5.991	0.270	4.609	7.469
Range	63.439	1.255	1.978	20.799	32.786	0.328	-858.333	985.210	1521.035	31.880	118.299
Mixed: pt & L/40											
Sill	6.131	0.442	7.203	0.799	13.027	0.130	5.790	6.472	3.431	4.809	7.951
Range	75.404	13.220	17.532	23.792	31.553	0.316	-972.590	1123.398	2076.800	37.100	125.334
Mixed: L/80 & L/40											
Sill	5.725	0.036	0.637	0.742	12.965	0.130	5.645	5.806	0.181	4.690	7.528
Range	91.915	29.731	32.346	18.737	20.385	0.204	-687.276	871.106	1329.773	56.296	135.466
pt + L/80											
Sill	5.994	0.305	5.094	0.894	14.915	0.149	5.880	6.108	0.227	4.517	8.511
Range	55.394	6.790	12.258	29.606	53.447	0.534	-394.454	505.242	1098.842	16.136	120.119
pt + L/40											
Sill	6.244	0.555	8.887	0.914	14.633	0.146	6.111	6.377	0.302	4.943	8.620
Range	79.643	17.459	21.922	37.324	46.864	0.469	-595.565	754.852	2048.977	28.692	156.471
L/80 + L/40											
Sill	6.076	0.387	6.365	0.821	13.519	0.135	5.994	6.158	0.157	4.728	8.217
Range	73.109	10.925	14.943	32.335	44.229	0.442	-316.712	462.930	978.506	22.286	138.678
All variograms											
Sill	6.032	0.343	5.694	0.791	13.113	0.131	6.003	6.061	0.055	4.756	7.903
Range	70.240	8.056	11.470	25.128	35.775	0.358	-58.214	198.694	226.830	22.962	115.208

Table E.2: Experiment 3 Nugget, 12x12, Sill = 5.689, Range = 3*62.184. Confidence Intervals (C.I.) are at the 95% level

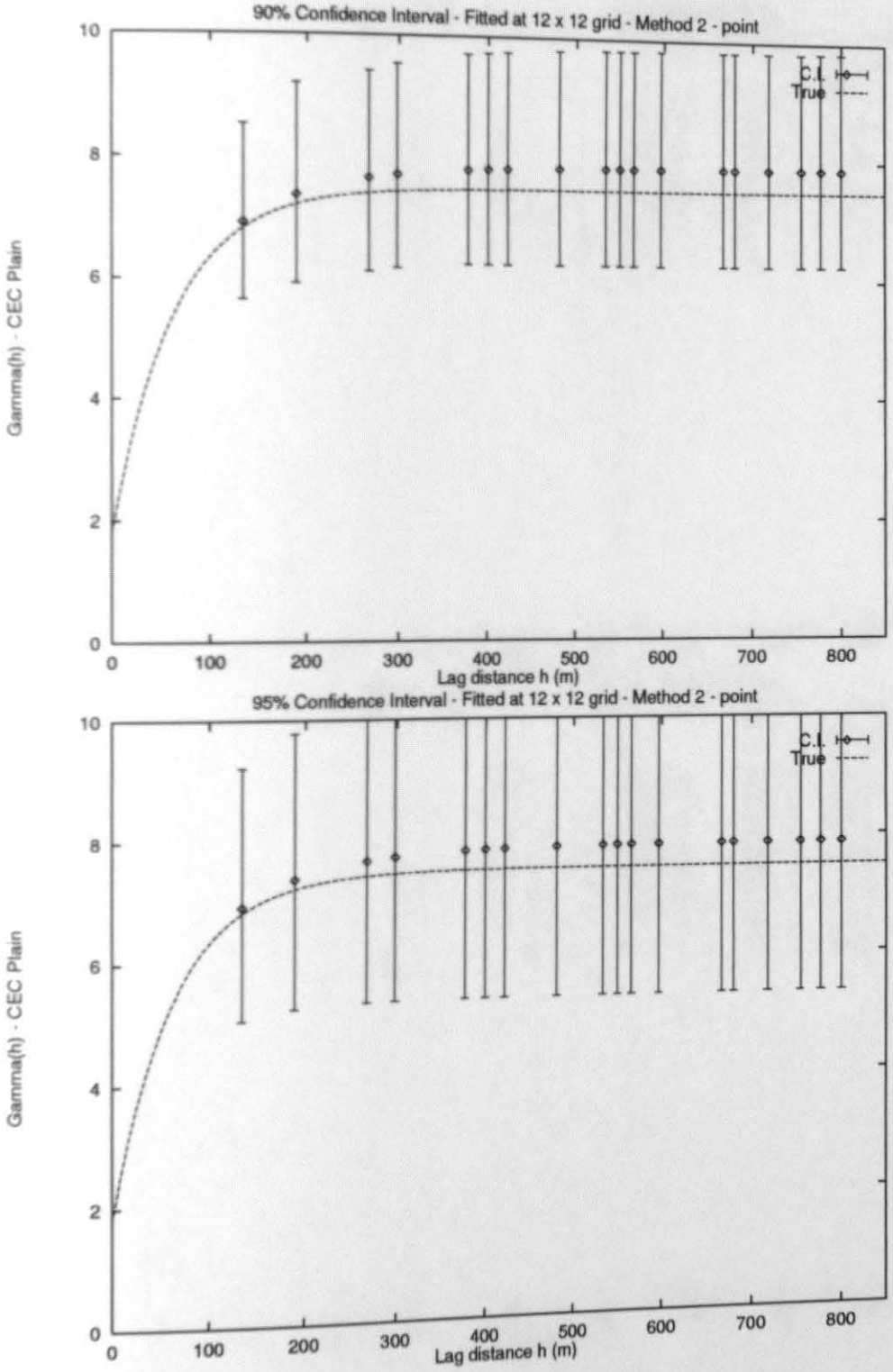


Figure E.11: Numerical Experiment 3 Nugget, (192 points) - point

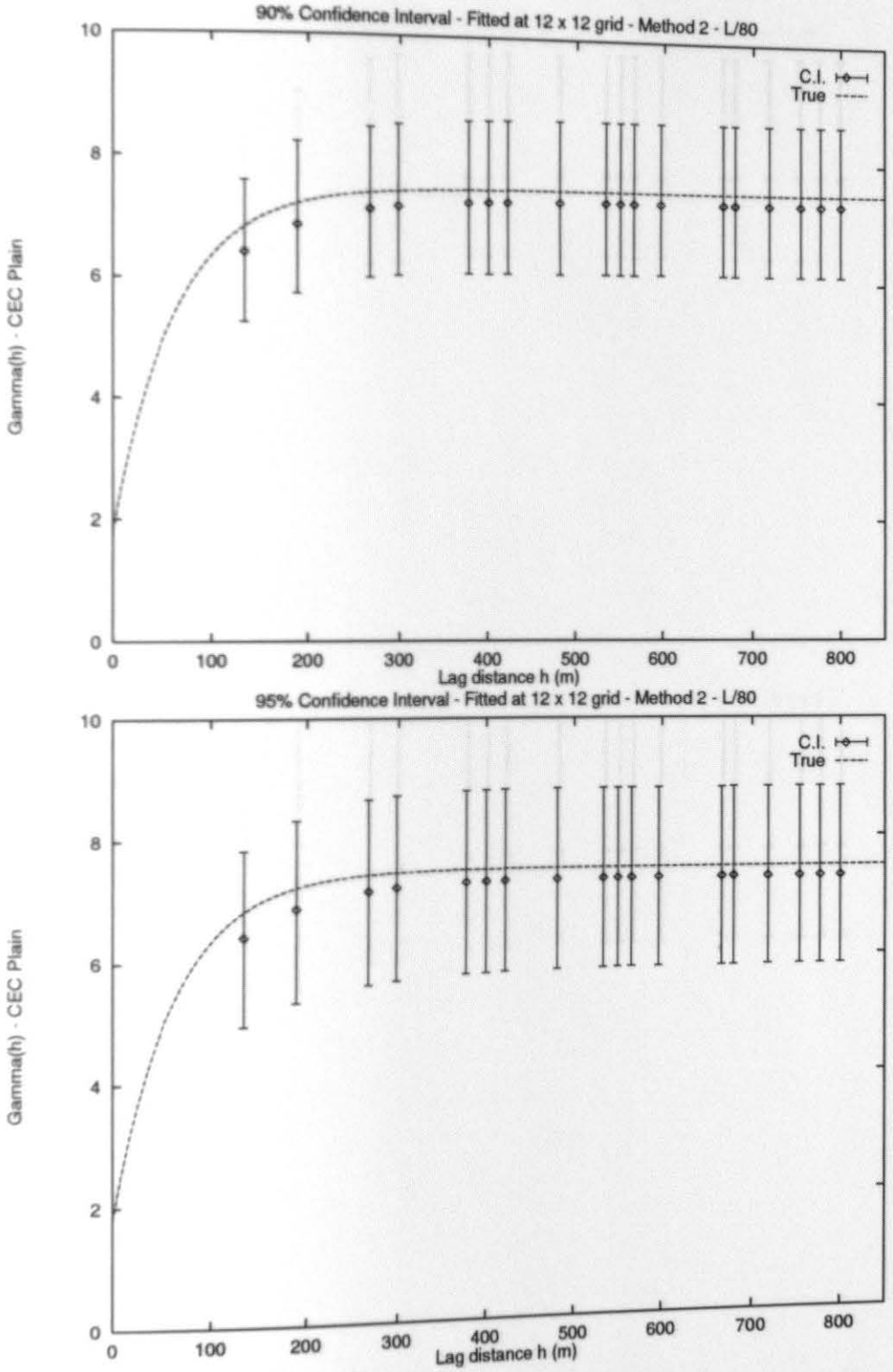


Figure E.12: Numerical Experiment 3 Nugget, (192 points) - L/80

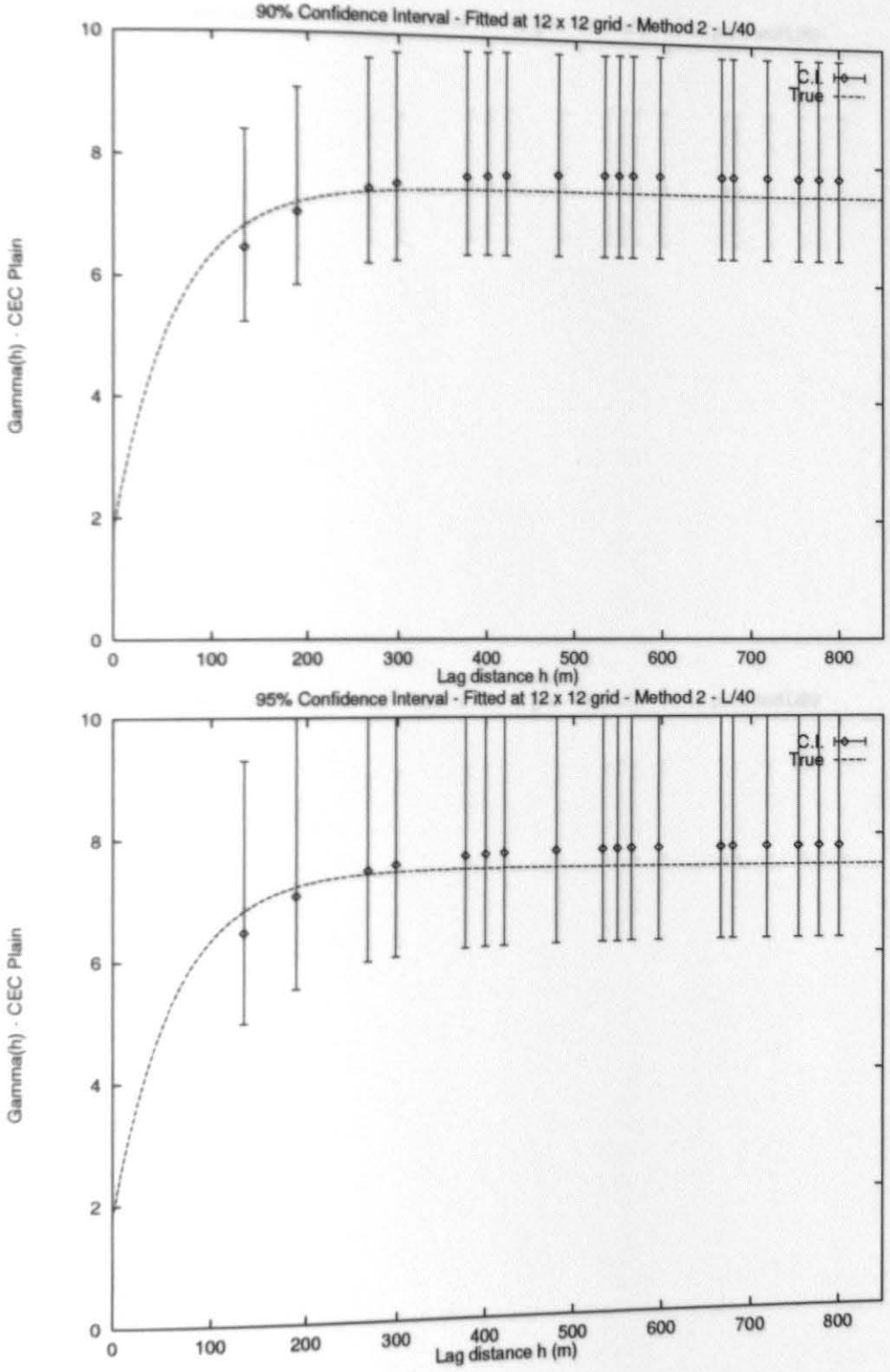


Figure E.13: Numerical Experiment 3 Nugget, (192 points) - L/40

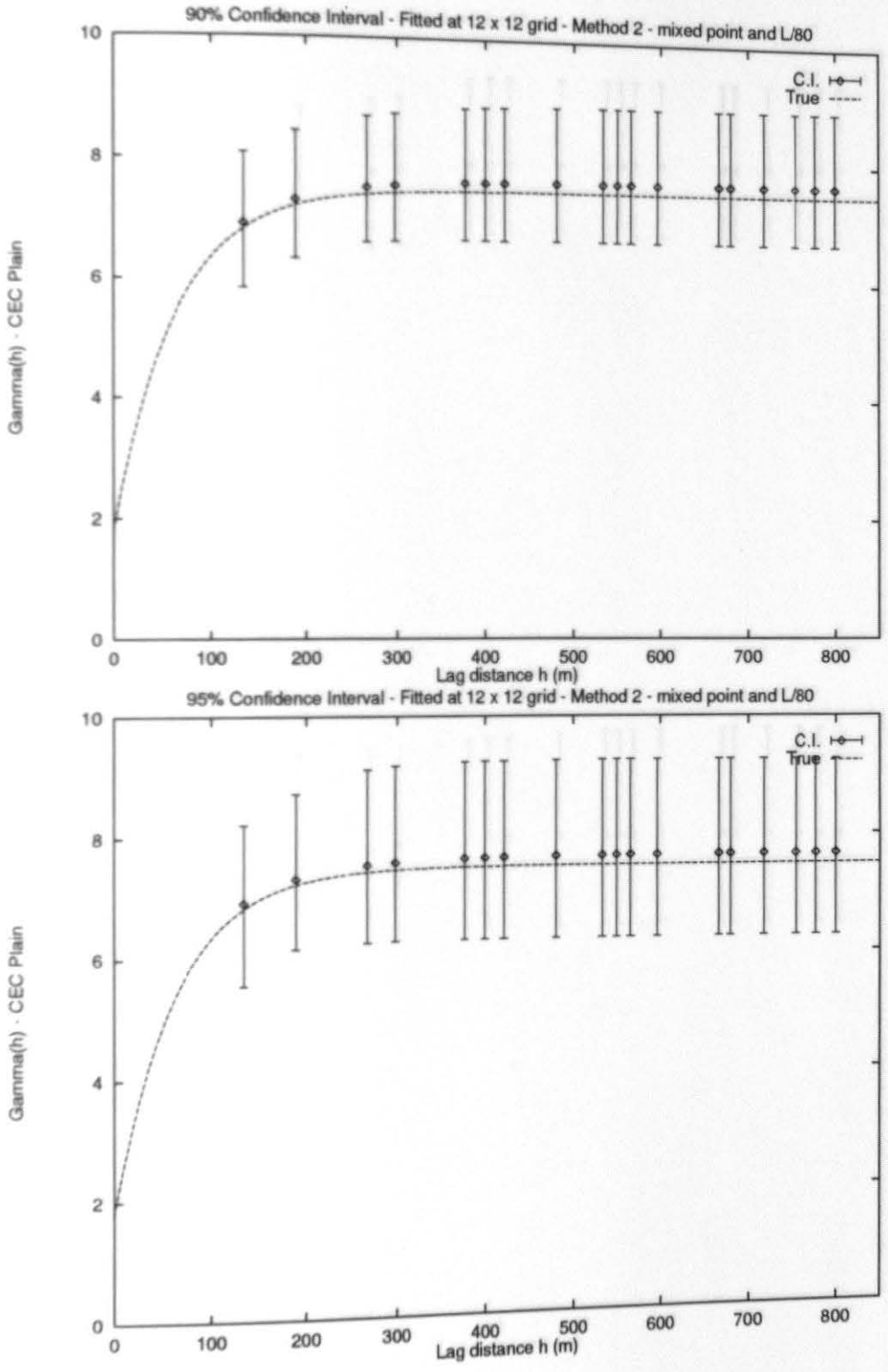


Figure E.14: Numerical Experiment 3 Nugget, (192 points) - mixed point and L/80

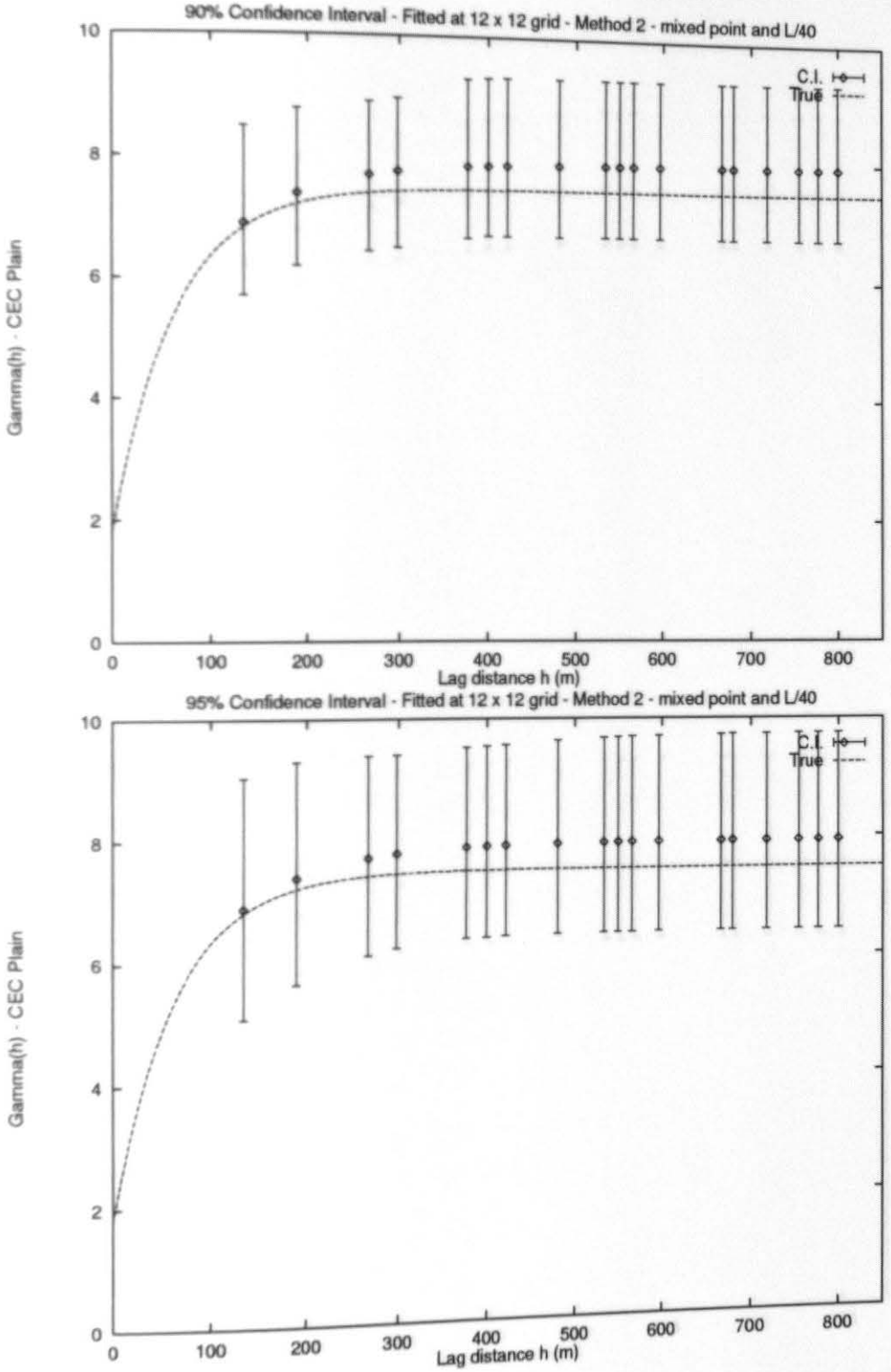


Figure E.15: Numerical Experiment 3 Nugget, (192 points) - mixed point and L/40

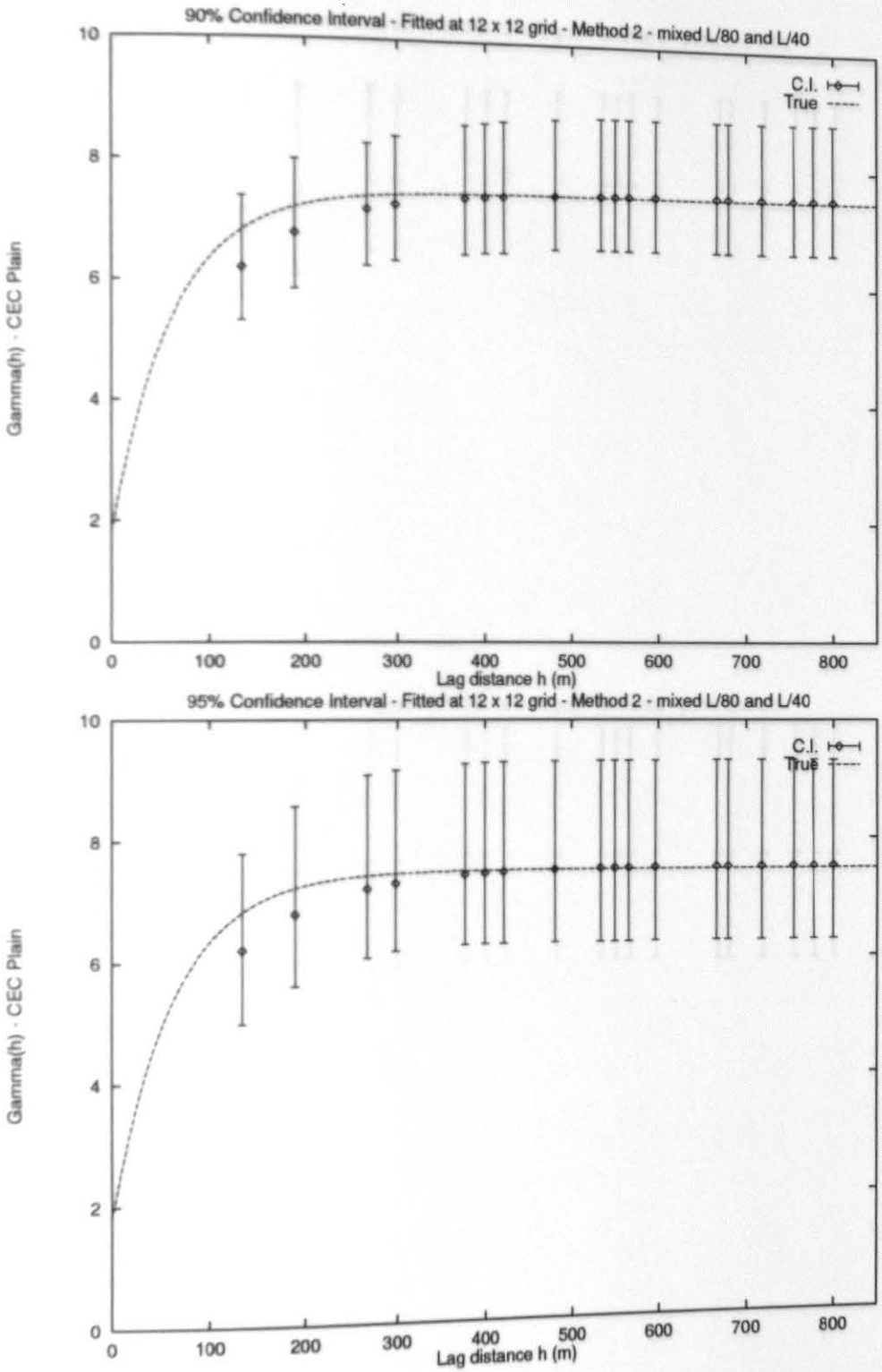


Figure E.16: Numerical Experiment 3 Nugget, (192 points) - mixed L/80 and L/40

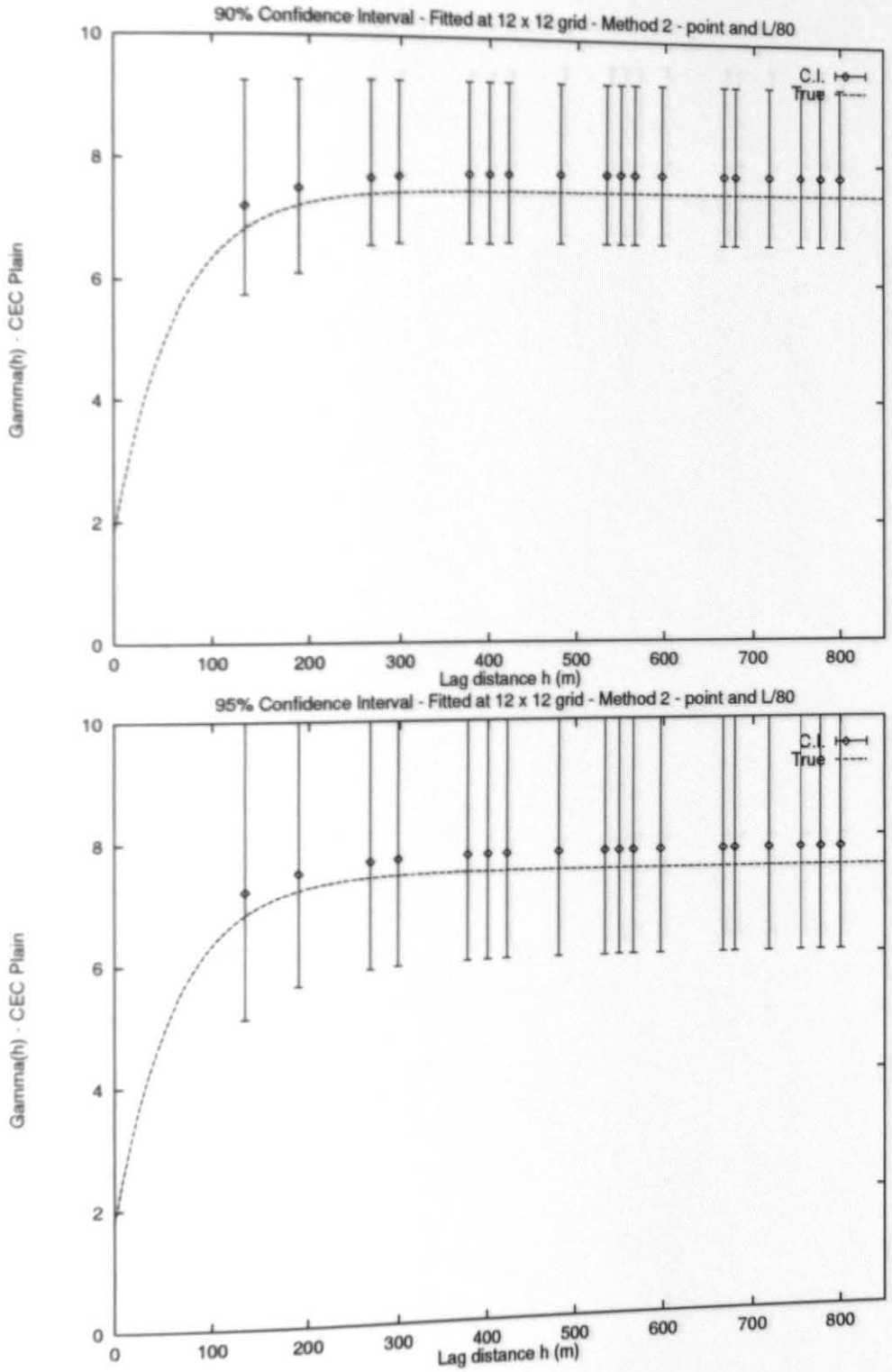


Figure E.17: Numerical Experiment 3 Nugget, (192 points) - point and L/80

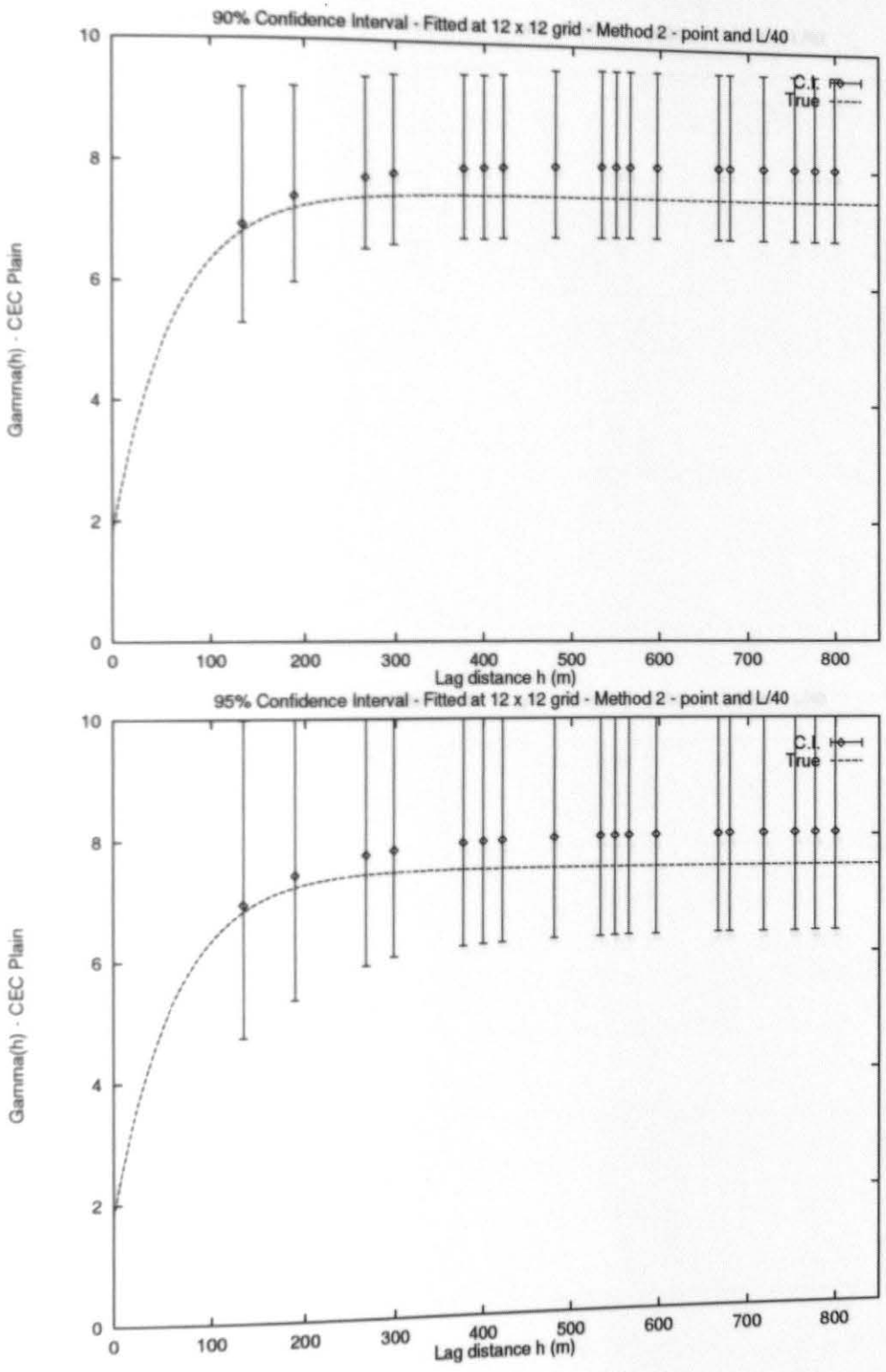


Figure E.18: Numerical Experiment 3 Nugget, (192 points) - point and L/40

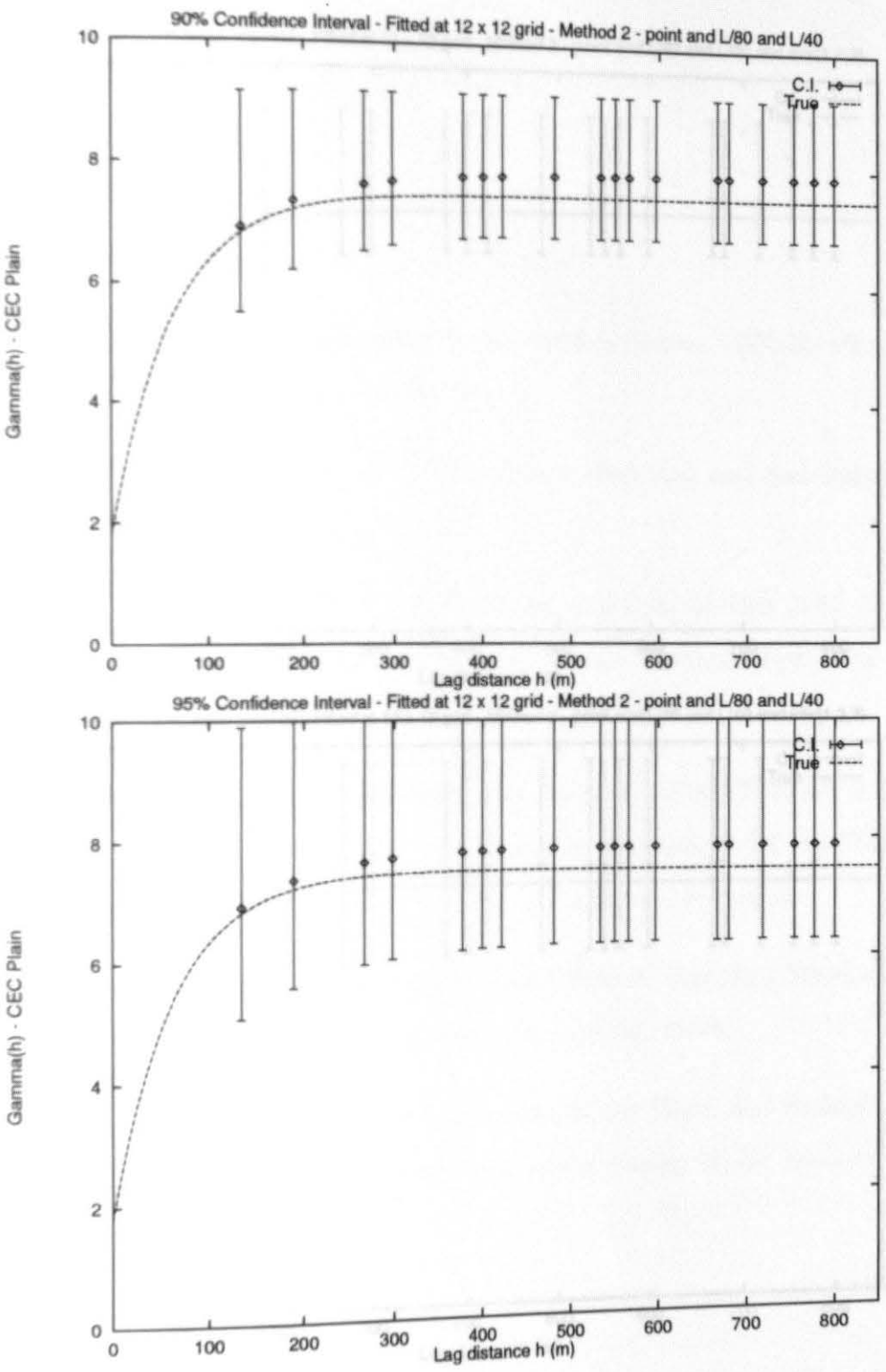


Figure E.19: Numerical Experiment 3 Nugget, (192 points) - point and L/80 and L/40

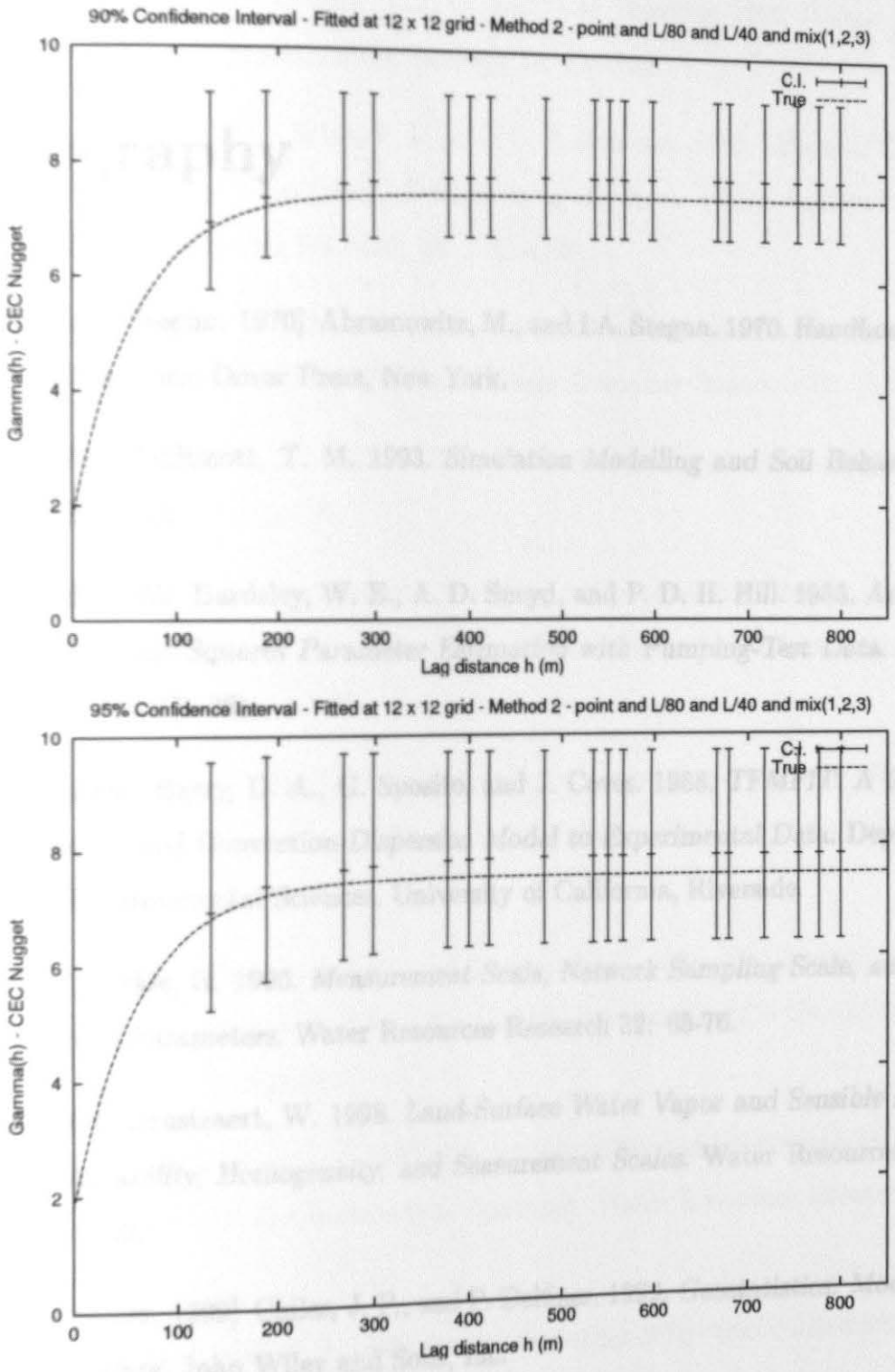


Figure E.20: Numerical Experiment 3 Nugget, (192 points) - all variograms at once

Bibliography

[Abramowitz and Stegun, 1970] Abramowitz, M., and I.A. Stegun. 1970. *Handbook of mathematical functions.* Dover Press, New York.

[Addiscott, 1993] Addiscott, T. M. 1993. *Simulation Modelling and Soil Behaviour.* Geoderma 60: 15-40.

[Bardsley et al., 1985] Bardsley, W. E., A. D. Sneyd, and P. D. H. Hill. 1985. *An Improved Method of Least-Squares Parameter Estimation with Pumping-Test Data.* Journal of Hydrology 80: 271-281.

[Barry et al., 1988] Barry, D. A., G. Sposito, and J. Coves. 1988. *TFMFIT: A Program to Fit a Generalized Convection-Dispersion Model to Experimental Data.* Department of Soil and Environmental Sciences. University of California, Riverside.

[Beckie, 1996] Beckie, R. 1996. *Measurement Scale, Network Sampling Scale, and Groundwater Model Parameters.* Water Resources Research 32: 65-76.

[Brutsaert, 1998] Brustsaert, W. 1998. *Land-Surface Water Vapor and Sensible Heat Flux: Spatial Variability, Homogeneity, and Seasurement Scales.* Water Resources Research 34: 2433-2442.

[Chiles and Delfiner, 1999] Chiles, J. P., and P. Delfiner. 1999. *Geostatistics: Modeling Spatial Uncertainty.* John Wiley and Sons, Inc.

[Cressie, 1985] Cressie, N. A. C. 1985. *Fitting Variogram Models by Weighted Least Squares.* Mathematical Geology 17: 563-586.

[Cressie, 1993] Cressie, N. A. C. 1993. *Statistics for Spatial Data.* John Wiley and Sons, Inc.

[Cressie and Hawkins, 1980] Cressie, N. A. C., and D. M. Hawkins. 1980. *Robust Estimation of the Variogram: I.* Mathematical Geology 12: 115-125.

[Clifton and Neuman, 1982] Clifton, P. M., and S. P. Neuman. 1982. *Effects of Kriging and Inverse Modeling on Conditional Simulation of the Avra Valley Aquifer in Southern Arizona* Water Resources Research 18: 1215-1234.

[Cooley, 1985] Cooley, R. L. 1985. *A Comparison of Several Methods of Solving Nonlinear Regression Groundwater Flow Problems.* Water Resources Research 21: 1525-1538.

[Cushman, 1986] Cushman, J. H. 1986. *On Measurement, Scale, and Scaling.* Water Resources Research 22: 129-134.

[Dagan, 1982] Dagan, G. 1982. *Stochastic Modeling of Groundwater Flow by Unconditional and Conditional Probabilities: 1. Conditional Simulation and the Direct Problem.* Water Resources Research 18: 813-833.

[Dagan, 1985] Dagan, G. 1985. *Stochastic Modeling of Groundwater Flow by Unconditional and Conditional Probabilities: The Inverse Problem.* Water Resources Research 21: 65-72.

[Delay and de Marsily, 1994] Delay, F., and De Marsily, G. 1994. *The Integral of the Semivariogram: A Powerful Method for Adjusting the Semivariogram in Geostatistics.* Mathematical Geology 26: 301-321.

[Delhomme, 1979] Delhomme, J. P. 1979. *Spatial Variability and Uncertainty in Groundwater Flow Parameters: A Geostatistical Approach.* Water Resources Research 15: 269-280.

[Desbarats, 1994] Desbarats, A. J. 1994. *Spatial Averaging of Hydraulic Conductivity Under Radial Flow Conditions.* Mathematical Geology 26: 1-21.

[Deutsch and Journel, 1997] Deutsch, C. V., and A. G. Journel. 1997. *GSLIB: Geostatistical Software Library and User's Guide.* Second Edition. Oxford University Press, New York.

[Ellsworth, 1996] Ellsworth, T.R. 1996. *Influence of Transport Variability Structure on Parameter Estimation and Model Discrimination.* Application of GIS to the Modeling of Non-Point Source Pollutants in the Vadose Zone, Soil Sci. Soc. Am. Special Publication 48.

[Ellsworth et al., 1996] Ellsworth, T.R., P.J. Shouse, T.H. Skaggs, J.A. Jobes and J. Fargerlund. 1996. *Solute transport in unsaturated soil: Experimental design, parameter estimation and model discrimination.* Soil Sci. Soc. Am. J. 60: 397-407.

[Ellsworth and Boast, 1996] Ellsworth, T.R., and C.W. Boast. 1996. *Spatial structure of solute transport variability in an unsaturated field soil.* Soil Sci. Soc. Am. J. 60: 1355-1367.

[Gelhar, 1993] Gelhar, L. W. 1993. *Stochastic Subsurface Hydrology.* Prentice-Hall, Inc. Englewood Cliffs, New Jersey.

[Gómez-Hernández and Gorelick, 1989] Gómez-Hernández, J. J., and S. Gorelick. 1989. *Effective Groundwater Model Parameter Values: Influence of Spatial Variability of Hydraulic Conductivity, Leakance, and Recharge.* Water Resources Research 25: 405-419.

[Goovaerts, 1993] Goovaerts, P. 1993. *Spatial Orthogonality of the Principal Components Computed from Coregionalized Variables.* Mathematical Geology 25: 281-302.

[Goovaerts, 1997] Goovaerts, P. 1997. *Geostatistics for Natural Resources Evaluation.* Oxford University Press, New York.

[Gotway and Hergert, 1997] Gotway, C. A., and G. W. Hergert. 1997. *Division S-8 - Nutrient Management & Soil & Plant Analysis.* Soil Sci. Soc. Am. J. 61: 298-309.

[Graham and McLaughlin, 1991] Graham, W. D., and D. B. McLaughlin. 1991. *A Stochastic Model of Solute Transport in Groundwater: Application to the Borden, Ontario, Tracer Test.* Water Resources Research 27: 1345-1359.

[Hawkins and Cressie, 1984] Hawkins, D. M., and Cressie, N. A. C. 1984. *Robust Kriging - A Proposal.* Mathematical Geology 16: 3-18.

[Heuvelink, 1998] Heuvelink, G. B. M. 1998. *Uncertainty Analysis in Environmental Modelling Under a Change of Spatial Scale.* Nutrient Cycling in Agroecosystems 50: 255-264.

[Hoeksema and Kitanidis, 1985a] Hoeksema, R. J., and P. K. Kitanidis. 1985. *Analysis of the Spatial Structure of Properties of Selected Aquifers.* Water Resources Research 21: 563-572.

[Hoeksema and Kitanidis, 1985b] Hoeksema, R. J., and P. K. Kitanidis. 1985. *Comparison of Gaussian Conditional Mean and Kriging Estimation in the Geostatistical Solution of the Inverse Problem.* Water Resources Research 21: 825-836.

[Isaaks and Srivastava, 1989] Isaaks, E. H., and R. M. Srivastava. 1989 *An Introduction to Applied Geostatistics.* Oxford University Press, New York.

[Journel and Huijbregts, 1978] Journel, A.G., and Ch.J. Huijbrechts. 1978. *Mining Geostatistics.* Academic Press, London.

[Jury and Sposito, 1985] Jury, W. A., and G. Sposito. 1985 *Field Calibration and Validation of Solute Transport Models for the Unsaturated Zone* Soil Sci. Soc. Am. J. 49: 1331-1341.

[Kitanidis, 1995] Kitanidis, P. K. 1995 *Quasi-Linear Geostatistical Theory For Inversing* Water Resources Research 31: 2411-2419.

[Kitanidis and Lane, 1985] Kitanidis, P. K., and R. B. Lane. 1985 *Maximum Likelihood Parameter Estimation of Hydrologic Spatial Processes By The Gauss-Newton Method* Journal of Hydrology 79: 53-71.

[Kitanidis and Vomvoris, 1983] Kitanidis, P. K., and E. G. Vomvoris. 1983 *A Geostatistical Approach to the Inverse Problem In Groundwater Modeling (Steady State) and One-Dimensional Simulations* Water Resources Research 19: 677-690.

[Knopman and Voss, 1987] Knopman, D. S., and C. I. Voss. 1987 *Behavior of Sensitivities in the One-Dimensional Advection-Dispersion Equation: Implications for Parameter Estimation and Sampling Design* Water Resources Research 23: 253-272.

[Kopka and Daly, 1997] Kopka, H. and P.W. Daly. 1997. *A Guide to LaTeX 2ε, Second Edition*. Addison-Wesley Publishers Ltd.

[Lamorey and Jacobson, 1995] Lamorey, G. and E. Jacobson. 1995. *Estimation of Semivariogram Parameters and Evaluation of the Effects of Data Sparsity*. Mathematical Geology 27: 327-358.

[Matheron, 1989] Matheron, G. 1989. *Estimating and Choosing*. Springer-Verlag, New York Inc.

[McLaughlin, 1995] McLaughlin, D. 1995. *Recent Developments in Hydrologic Data Assimilation*. Reviews of Geophysics, Supplement, pp. 977-984. American Geophysical Union.

[McLaughlin and Townley, 1996] McLaughlin, D. and L. R. Townley. 1996. *A Reassessment Of The Groundwater Inverse Problem*. Water Resources Research 32: 1131-1161.

[McLaughlin and Wood, 1988] McLaughlin, D. and E. F. Wood. 1988. *A Distributed Parameter Approach for Evaluating the Accuracy of Groundwater Model Predictions: 1. Theory*. Water Resources Research 24: 1037-1047.

[Murty and Scott, 1977] Murty, V. V. N., and V. H. Scott. 1977. *Determination of Transport Model Parameters in Groundwater Aquifers*. Water Resources Research 13: 941-947.

[Neuman, 1990] Neuman, S. P. 1990. *Universal Scaling of Hydraulic Conductivities and Dispersivities in Geologic Media*. Water Resources Research 26: 1749-1758.

[Parker and van Genuchten, 1984] Parker, J. C. and M. Th. van Genuchten. 1984. *Determining Transport Parameters from Laboratory and Field Tracer Experiments*. Bulletin 84-3, 1984. Virginia Agricultural Experiment Station. Virginia Polytechnic Institute and State University.

[Parker and Albrecht, 1987] Parker, J. C. and K. A. Albrecht. 1987. *Sample Volume Effects on Solute Transport Predictions*. Water Resources Research 23: 2293-2301.

[Petach et al., 1991] Petach, M. C., R. J. Wagenet, and S. D. DeGloria. 1991. *Regional Water Flow and Pesticide Leaching Using Simulations with Spatially Distributed Data.* Geoderma 48: 245-269.

[Press et al., 1992] Press, W.H., B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. 1992. *Numerical Recipes in C: The Art of Scientific Computing, Second Edition.* Cambridge University Press, New York.

[Rajaram and McLaughlin, 1990] Rajaram, H., and D. McLaughlin. 1990. *Identification of Large-Scale Spatial Trends in Hydrologic Data.* Water Resources Research 26: 2411-2423.

[Rouhani, 1985] Rouhani, S. 1985. *Variance Reduction Analysis.* Water Resources Research 21: 837-846.

[Russo, 1984] Russo, D. 1984 *Design of an Optimal Sampling Network for Estimating the Variogram.* Soil Sci. Soc. Am. J. 48: 708-716.

[Russo and Jury, 1987a] Russo, D., and W. A. Jury. 1987 *A Theoretical Study of the Estimation of the Correlation Scale in Spatially Variable Fields. 1. Stationary Fields.* Water Resources Research 23: 1257-1268.

[Russo and Jury, 1987b] Russo, D., and W. A. Jury. 1987 *A Theoretical Study of the Estimation of the Correlation Scale in Spatially Variable Fields. 2. Nonstationary Fields.* Water Resources Research 23: 1269-1279.

[Sadeghipour and Yeh, 1984] Sadeghipour, J. and W. W-G. Yeh. 1984 *Parameter Identification of Groundwater Aquifer Models: A Generalized Least Squares Approach.* Water Resources Research 20: 971-979.

[Sánchez-Vila et al., 1995] Sánchez-Vila, X., J. P. Girardi, and J. Carrera. 1995 *A Synthesis of Approaches to Upscaling of Hydraulic Conductivities.* Water Resources Research 31: 867-882.

[Schwartz and Christiansen, 1997] Schwartz, R. L., and T. Christiansen. 1997 *Learning Perl, Second Edition*. O'Reilly and Associates, Inc., Sebastopol, California.

[Schildt, 1996] Schildt, H.P. 1996. *The Complete C++ Reference*. Osborne-McGraw Hill, New York.

[Shafer and Varljen, 1990] Shafer, J. M., and M. D. Varljen. 1990. *Approximation of Confidence Limits on Sample Semivariograms From Single Realizations of Spatially Correlated Random Fields*. Water Resouces Research 26: 1787-1802.

[Shouse et al., 1994] Shouse, P. J., T. R. Ellsworth, and J. A. Jobes. 1994. *Steady-state infiltration as a function of measurement scale*. Soil Sci. Soc. Am. J. 157: 129-136.

[Sorooshian and Gupta, 1985] Sorooshian, S., and V. K. Gupta. 1985. 1985. *The Analysis of Structural Identifiability: Theory and Application to Conceptual Rainfall-Runoff Models*. Water Resources Research 21: 487-495.

[Stein et al., 1988] Stein, A., M. Hoogerwerf, and J. Bouma. 1988. *Use of Soil-Map Delineations to Improve (Co-)Kriging of Point Data on Moisture Deficits*. Geoderma 43: 163-177.

[Sun, 1994] Sun, N.-Z. 1994. *Inverse Problems in Groundwater Modeling*. Kluwer Academic, Norwell, Massachussetts.

[Tabor et al., 1985] Tabor, J. A., A. W. Warrick, D. E. Myers, and D. A. Pennington 1985. *Spatial Variability of Nitrate in Irrigated Cotton: II. Soil Nitrate and Correlated Variables*. Soil Sci. Soc. Am. J. 49: 390-394.

[Tercan and Dowd, 1993] Tercan, A. E., and P. A. Dowd 1993. *Approximate Local Confidence Intervals Under Change of Support*. Mathematical Geology 25: 149-172.

[Thompson and Seber, 1996] Thompson, S. K., and G. A. F. Seber 1996. *Adaptive Sampling*. Wiley Series in Probability and Statistics, New York, New York.

[Tiktak et al., 1998] Tiktak, A., J. R. M. Alkemade, J. J. M. van Grinsven, and G. B. Makaske. 1998. *Modelling Cadmium Accumulation at a Regional Scale in the Netherlands.* Nutrient Cycling in Agroecosystems 50: 209-222.

[Wagner and Gorelick, 1986] Wagner, B. J., and S. M. Gorelick. 1986. *A Statistical Methodology for Estimating Transport Parameters: Theory and Applications to One-Dimensional Advective-Dispersive Systems.* Water Resources Research 22: 1303-1315.

[Warrick and Myers, 1987] Warrick, A. W., and D. E. Myers. 1987. *Optimization of Sampling Locations for Variogram Calculations.* Water Resources Research 23: 496-500.

[Warrick et al., 1999] Warrick, A. W., J. A. Vargas-Guzmán, S. Mayer, M. Restrepo, and T. R. Ellsworth. 1999. *Strategies for Improving Spatial Variability Assessments.* Assessment of Non-Point Source Pollution in the Vadose Zone, Geophysical Monograph 108, American Geophysical Union.

[Webster, 1985] Webster, R. 1985. *Advances in Soil Science, Volume 3.* Springer-Verlag, New York Inc.

[Yeh, 1986] Yeh, W. W.-G. 1986. *Review of Parameter Identification Procedures in Groundwater Hydrology: The Inverse Problem.* Water Resources Research 22: 95-108.

[Yeh and Yoon, 1981] Yeh, W. W.-G., and Y. S. Yoon. 1981. *Aquifer Parameter Identification with Optimum Dimension in Parametrization.* Water Resources Research 17: 664-672.

[Zhang et al., 1990] Zhang, R., A. W. Warrick, and D. E. Myers. 1990. *Variance as a Function of Sample Support.* Mathematical Geology 22: 107-121.

Vita

Miguel Restrepo was born in Barranquilla, Colombia on May 14, 1968. He is the son of Guillermo (b.1935) and Sylvia (b. 1941) Restrepo. He has two older brothers, Esteban (b. 1963) and Andres (b. 1965). When he was four months old, the entire family moved to Bogotá. At age 4, he started from Kindergarten at the Gimnasio Moderno school. He studied at the Gimnasio Moderno for all the Primary, Middle, Junior High, and High School. He graduated on November, 1985 with the degree of "Bachiller Academico", being the top student of his class. In January, 1986 he joined the undergraduate program in Civil Engineering at the Universidad de Los Andes in Bogotá. The undergraduate program in Colombia lasts 5 years and requires the completion of a senior thesis. He minored in mathematics, and on 1991, being one semester short of graduation, he participated for one semester in the undergraduate exchange program with the University of Illinois. He spent the Spring of 1991 in Urbana, where he enjoyed the classes that he took: ocean structures, structural dynamics, abstract algebra, real analysis, and differential geometry. When he came back to Colombia, he finished his senior thesis and graduated on September 7, 1991 with his B.Sc. degree. His thesis was on the area of geophysics, with an emphasis of numerical analysis to simulate electric resistivity prospecting. At the time of his graduation, he was working at the "Laboratorio de Ensayos Hidraulicos" which was managed by a company called "Hidroestudios S.A.". He worked there for 6 months and the worked at his father's company, "Restrepo y Uribe Ltda." for another 6 months. After that, in August, 1992, he came to the United States to join the Master of Engineering program in Water Resources at Cornell University. The program there is brutal, requiring six classes per semester in order to complete the graduation requirements during the one year that the program is required to last. He

worked with Professor Christine Shoemaker and under the advisory of Professor Daniel P. Loucks as well. At the end of the program, he was accepted to the graduate program in Environmental Engineering at the University of Illinois at Urbana-Champaign. He started during August, 1993, working for professor Albert J. Valocchi and Wayland Eheart. He finished the Master of Science degree and graduated in August, 1996, and he joined then the newly formed department of Natural Resources and Environmental Sciences at the same university. He worked under the supervision of Professor Tim Ellsworth in the area of Soil Physics and Geostatistics. After assisting to several conferences and presenting posters and talks, he presented his prelim exam on October 18, 1998. Having successfully passed it, he started working on his dissertation. His dissertation work took about two years, and it was entitled "Variability Structure Estimation: The Role of Sample Support". During his dissertation work, he also worked as a research assistant at the Illinois State Water Survey, under the supervision of Dr. David Soong and Mrs. Laura Keefer. He worked then in several projects regarding the River Illinois and also several minor streams and lakes. During the month of July, he will start working at the EROS Data Center with Raytheon Corp. in Sioux Falls, South Dakota.